PROCRASTINATION IN THE WORKPLACE: EVIDENCE FROM THE U.S. PATENT OFFICE*

MICHAEL D. FRAKES AND MELISSA F. WASSERMAN

Despite much theoretical attention to the concept of procrastination and much exploration of this phenomenon in laboratory settings, there remain few empirical investigations into the practice of procrastination in real world contexts, especially in the workplace. In this paper, we attempt to fill these gaps by exploring procrastination among U.S. patent examiners. We find that nearly half of examiners' first substantive reports are completed immediately prior to the operable deadlines. Moreover, we find a range of additional empirical markers to support that this "end-loading" of reviews results from a model of procrastination rather than various alternative time-consistent models of behavior. In one such approach, we take advantage of the natural experiment afforded by the Patent Office's staggered implementation of its telecommuting program, a large-scale development that we theorize might exacerbate employee self-control problems due to the ensuing reduction in direct supervision. Supporting the procrastination theory, we estimate an immediate spike in application end-loading and other indicia of procrastination upon the onset of telecommuting. Finally, contributing to a growing empirical literature over the efficiency of the patent examination process, we assess the consequences of procrastination for the quality of the reviews completed by the affected examiners. This analysis suggests that the primary harm stemming from procrastination is delay in the ultimate application process, with rushed reviews completed at deadlines resulting in the need for revisions in subsequent rounds of review. Our findings imply that nearly 1/6 of the annual growth in the Agency's much-publicized backlog may be attributable to examiner procrastination.

JEL Codes: *D03, J45, M50, O34*

---

* Frakes (corresponding author): Duke University School of Law, NBER, michael.frakes@law.duke.edu, 919-613-7185, 210 Science Drive, Box 90360, Durham, NC 27708-0360; Wasserman: University of Texas School of Law. The work was funded by University of Illinois at Urbana Champaign Research Board, Award 12088 and the Cornell Institute for the Social Sciences Small Grant Award. We are grateful to Matt Berry at the National Center for Supercomputing Applications at the University of Illinois for collecting data from the Patent Office's PAIR database. We are likewise grateful to Jacob Goldin, Mark Lemley, Paul Oyer, Arti Rai, Kyle Rozema, Bhaven Sampat, various anonymous patent examiners, and seminar participants at the USPTO Empirical Patent Law Roundtable; the Annual Intellectual Property Scholars Conference; the Northwestern Law and Economics Colloquium; the Duke Law, Economics and Politics Workshop; the University of Texas Law Faculty Drawing Board, and the Conference on Empirical Legal Studies for helpful comments.

I. INTRODUCTION

A large body of literature has theorized that economic actors may counterproductively delay—i.e., procrastinate—on tasks that they had previously agreed to undertake. This lapse in self-control is often thought to stem from models of individual decisionmaking that involve either salience costs—e.g., differential discount rates applied to the costs and benefits of tasks—or present-biased preferences—e.g., higher discount rates applied to the short term versus the long term (Akerlof 1991; O'Donoghue and Rabin 1999). Consider for instance, a scholar who agrees to review a paper for an academic journal. When she accepts the assignment, the benefits from the work may appear to outweigh the costs. However, as the time nears to complete the referee report, the costs take on an enhanced degree of saliency in the eyes of the reviewer, causing her to deviate from her initial plans to commence work (Ariely and Wertenbroch 2002).

While the concept of procrastination has been the subject of various empirical studies to date, this empirical literature faces several limitations. To begin, documentation of present bias and time-inconsistent behavior has largely focused on decisions that are more personal in nature. Very few studies have explored these phenomena in the workplace. Second, the literature has predominantly drawn its insights from laboratory settings. Surprisingly little work has been done to explore procrastination of real world tasks, especially those performed by high skilled laborers. Finally, and critically, few investigations into the presence of procrastination have addressed the impacts of such behavioral tendencies on the quality of the tasks ultimately performed.

In this paper, we begin to fill these gaps in the literature by exploring the behaviors of a particular set of high-skilled laborers: patent examiners at the United States Patent Office. Charged with assessing the patentability of claimed inventions, patent examiners perform tasks of substantial import to innovation policy and economic growth.[1] Commentators have expressed growing concerns that the Patent Office is failing to provide high quality review of patent applications, implicating significant social welfare harms (e.g., Lemley and Sampat 2012, Frakes and Wasserman 2017a, Frakes and Wasserman 2015). Among these concerns are the harms stemming from allegedly rampant examiner procrastination, a topic that has been the subject of two recent reports by the Office of Inspector General of the Department of Commerce (OIG). These reports provide some limited statistics demonstrating that patent examiners frequently "end-load" their reviews—i.e., submit a high volume of work product immediately before deadlines (OIG Report, 2015; OIG Report, 2014). The OIG reports surmised, without offering supporting analysis, that end-loading was caused by examiner procrastination and that this practice may be negatively impacting the work product of the Agency. In this paper, we build upon the OIG reports by systematically demonstrating the full extent to which examiners end-load their work efforts. Moreover, we move beyond the OIG reports by setting forth a range of additional findings that suggest that these practices are indeed a reflection of examiner procrastination and that speak to the effects of such procrastination on examination quality.

This attempt to document evidence of patent examiner procrastination confronts serious empirical challenges. First and foremost, finding evidence of employee procrastination requires information capable of establishing a proper benchmark—i.e., some sense of the timeline that a rational, time-consistent worker would follow in completing her work. Not only is it difficult to acquire systematic and easily quantifiable data on the work efforts of individual employees and on the timing of completion of the tasks that they perform (in addition to the quality of their performances), it is rare to find systematic information necessary to construct the no-

[1] On the more general topic of the significance of patent rights for the direction of economic growth, see Moser (2005).

procrastination counterfactual. In the present analysis, we establish this basis of comparison by taking advantage of the rigid quota system that examiners must follow every bi-week.

Though quotas themselves are devices that are implemented, in part, to curb procrastination tendencies, we embrace the implemented quota to test for some residual degree of procrastination among actual examiners. In essence, while a non-procrastinating examiner would be expected to spread her work equally (roughly) over the relevant period of time, a procrastinating examiner that is nonetheless motivated to hit her production target might tend to cluster her examinations at the end of the quota period. The degree to which examiners cluster their reviews in this manner may thus illuminate the degree of time inconsistency in examiner work effort.

Drawing on a novel dataset consisting of application-level data on nearly 2 million patent applications filed over a 10-year period, with information on the precise timing—to the day—of the numerous actions that examiners take, we follow this approach and find substantial evidence of examination bunching around quota-period ends. We focus our analysis on the key component of examiners' bi-weekly quota: the examiner's first substantive decision regarding the patentability of the claimed invention, known as the first office action on the merits. We find that roughly half of first office actions are completed and processed on the last day of the quota period.

We acknowledge that end-of-quota bunching of work product alone may accord with various time-consistent theories of behavior—e.g., application sorting—as distinct from true procrastination. We take a range of steps to help mediate the possible causes of this end-loading of examinations. For instance, as one might predict under a model of procrastination, we find evidence that examiners begin to end-load their applications at notably higher rates immediately upon their commencement of the Agency's much-publicized telecommuting program, a shift in their working environments that was implemented in a staggered manner and that removes one of the most common mechanisms to dealing with self-control problems among workers—i.e., in-person monitoring by supervisors.[2] In addition, and perhaps most importantly, we theorize various ways in which procrastination by examiners may impact the quality of their reviews—that is, we theorize outcomes that derive only from models of time inconsistency. We support a procrastination explanation for the observed end-loading of application reviews by, in turn, presenting evidence consistent with these theorized outcomes.

Critical to this examination quality analysis is the prediction that applications being processed at quota ends—given the limited number of hours in a day—will receive less attention per application than those processed in the interim periods. In the face of this rush, we theorize that examiners will proffer a quick and low-quality—a.k.a., "shotgun"—rejection on the first office action, regardless of the underling validity of the application. Examiners may opt for a quick rejection over a quick, ill-conceived allowance considering that first-office-action rejections are non-final in nature. That is, lacking the time to complete a thorough substantive review at the present, examiners may buy themselves time to provide this thorough review at a later date by rejecting the application now. In contrast, examiners do not have the chance to correct an improper allowance, an outcome that could compromise their performance reviews.

Consistent with this prediction, we find the rate of allowance on the first office action is drastically lower for those applications reviewed near the deadline. Moreover, our evidence suggests that applicants are able to overcome such "shotgun" rejections and that examiners are ultimately able to find time at a later date to review applications with nearly the same degree of scrutiny and care they apply to those applications initially reviewed in the middle of a quota period.

[2] Though we largely approach this telecommuting experiment in an effort to help assess the presence of self-control problems by patent examiners, this analysis also allows us to contribute to the growing literature on the productivity implications of telecommuting programs (Bloom et al. 2015).

Beyond supporting a time-inconsistent interpretation of the observed end-loading of reviews, this "shotgun"-rejection analysis is valuable insofar as we place independent importance on understanding the quality of the patent examination process. As such, our analysis not only begins to fill gaps in the behavioral and personnel economics literatures but also contributes to the nascent literature on the determinants of patent examiner behaviors and on the resulting welfare implications.[3] Ultimately, our analysis suggests that the immediate consequence of procrastination by examiners appears to be an increase in application processing time stemming from the need in subsequent rounds of review to make up for the "shotgun" rejections issued on end-loaded applications in earlier rounds. Examination delays may interfere with the deployment of valuable inventions to the marketplace and increase the uncertainty surrounding the rights of potential patents, which in turn may limit a company's ability to license or engage in related activity (Frakes and Wasserman, 2016). The Patent Office has repeatedly stated that its biggest challenge to fulfilling its mission of providing high quality timely reviews is its existing backlog of applications (Frakes and Wasserman 2015). Our estimates imply that the increases in processing delays stemming from procrastination of first office actions may have contributed to as much as 17% of the highly publicized growth rate in the Agency's backlog of applications over our sample period.

The paper proceeds as follows. In Part II, we review the existing literature on procrastination. In Part III, we provide a background on the patent examination process and theorize the ways in which non-procrastinating and procrastinating examiners approach the timing of their tasks. In Part IV, we discuss the data and methodologies that we employ to test the predictions from Part II and subsequently present the results of such tests. Finally, in Part V, we conclude.

[3] Contributions to this literature include Cockburn et al. 2003, Lemley and Sampat 2012, Frakes and Wasserman 2017a, and Frakes and Wasserman 2016. For a recent survey of this literature, Frakes and Wasserman 2017b.

II. LITERATURE REVIEW

Evidence bearing directly on the existence and degree of procrastination remains rather limited. More common are studies that have documented markers of time-inconsistent behavior and present bias more broadly, even if not specifically focused on delays in the completion of assigned tasks (of the sort that we commonly associate with the notion of procrastination). In this vein, scholars have documented evidence consistent with present bias in a such settings as caloric intake by food-stamp recipients (Shapiro 2005), tobacco consumption (Gruber and Koszegi 2001), gym membership (Acland and Levy 2015), life cycle savings (Laibson et al. 2007), food choice (Brown et al. 2009) and movie choices (Read, Lowenstein, and Kalyanaraman 1999).

As suggested by the topics of emphasis in these studies, the literature on present-bias has overwhelmingly focused on settings of a more personal nature. Rare are those studies focusing on procrastination or present-bias in the workplace, where the productivity consequences of these phenomena are potentially substantial. Insofar as the workplace generally incorporates external forms of control—e.g., supervision—and carries potentially significant consequences for poor performance—e.g., advancement, termination—there may be little reason to think the behavioral results from the non-workplace settings will generalize to this critical environment. One recent workplace investigation into self-control problems, however, is provided by Kaur, Kremer and Mullainathan (2015), which presents the results of a field experiment on data entry workers in India, finding evidence that when faced with the option of entering into a contract with piece-rate compensation versus a dominated contract that penalizes this compensation should workers fall below a target, many workers select the dominated option (presumably as a self-control device).[4]

A second limitation with the existing literature is that most of the supporting evidence of procrastination—and of time-inconsistent behaviors more generally—comes from the laboratory.[5] A small but growing number of studies have moved beyond the laboratory into a more natural, real-world setting by investigating these phenomena through field experiments.[6] We are aware of very few studies, if any, that have employed certain methodological techniques—including those quasi-experimental in nature—within an observational framework to document evidence suggestive of procrastination.[7] Though the methodological challenges facing observational approaches may be considerable, moving beyond a controlled experimental setting allows us to expand the scope of contexts in which we may explore these behaviors, especially into a high-skilled work environment where opportunities for experimental approaches may be more limited.

While an analysis of time-inconsistent behavior is interesting insofar as it challenges assumptions of rationality, one's primary interests arguably lie with the welfare implications of any such irrationalities. A final limitation of the literature is that the impact of procrastination on the quality of tasks performed is in need of greater study, having only been addressed in a limited number of studies to date (Cadena et. al, 2011; Kaur, Kremer and Mullainathan 2015).

[4] Other recent examples can be found in Cadena et al.'s (2011) analysis of the effectiveness of reminders in reducing deadline clustering in loan officers' work efforts and Duflo, Kremer and Robinson's (2011) analysis of the timing of fertilizer purchase decisions and of the demand for pre-purchasing programs.

[5] For an overview of such studies, see Frederick, Loewenstein and O'Donoghue (2002) and Sprenger (2015). A recent example can be found in Bisin and Hyndman (2014), which experimentally documents a nearly 40% rate of present bias among college students asked to perform various assignments—e.g., alphabetical sorting of word lists—within a week's time.

[6] Examples include Cadena et al. (2011), Kaur, Kremer and Mullainathan (2015) and Duflo, Kremer and Robinson (2011).

[7] One possible exception is found in Asch (1990), which estimated an increase in military recruiter productivity throughout a roughly one-year monitoring period. Asch suggests that this pattern may be consistent with procrastination; however, she also indicates that it may reflect a story in which recruiters work early in the observation period to stock-pile potential recruits and then pull from this pile to varying degrees later on in strategic attempts to receive particular rewards.

Our research on the Patent Office's quota is also related to a strand of the personnel economics literature addressing the impact of quota-based bonus schemes facing salespersons on the timing of sales contracts (Oyer, 1998; Larkin 2014).[8] This literature also predicts a spike in output at deadlines, however, these spikes may arise from entirely different mechanisms. For instance, in a sales contract context, end-of-period bunching of sales could result from salespersons trying to hit quotas by using price manipulations to "pull in" some sales at the end of one period that would have otherwise occurred in the beginning of the next period. In this sales example, lighter work product earlier in the period need not arise solely from a model of time-inconsistent behavior and worker delay, but could also result from lower-than-expected sales demand—i.e., from factors outside of the workers' control. This latter mechanism is arguably less relevant in the patent context, given that the Agency's substantial backlog of applications ensures that examiners always have reviews at their disposal that they can undertake in an effort to hit their quota.

Balasubramanian et al. (2017) discuss a "pull-in" story similar to that of the sales quota context. Interestingly, their analysis likewise focuses on the patent application process, though it explores the behavior of applicants and their attorneys, as opposed to examiners. They find that application filings by applicants tend to experience modest spikes on the last day of the month. Interviews with applicant attorneys reveal that these spikes may be due to monthly deadlines facing attorneys in many firms.[9] The authors, however, note that they are unable to distinguish a procrastination explanation for these spikes from a story in which applicant attorneys perhaps work longer than expected on applications earlier in the month and then accelerate work effort at the end of the month in order to meet the deadline (pulling applications that would have otherwise been filed

[8] For a related discussion, see Courty and Marshke 2004. For other research on the effects of deadlines, see Chetty et al. (2014).
[9] The nature of these applicant-related deadlines and the extent to which they apply to all applications (as is the case with the examiner deadlines we study) is unclear, however.

later). Interviews with applicant attorneys suggest that attorneys are more likely following the latter model. In our analysis of patent examiner behavior below, we attempt to separate an acceleration story from a delay story by employing tests that target self-control mechanisms—e.g., through the tele-commuting quasi-experiment discussed below.

III. BACKGROUND AND THEORY

III.A. Description of Examination Process

Every patent application filed with the Patent Office contains a specification, which describes the invention, and a set of claims that defines the metes and bounds of the legal rights the applicant is seeking. Before an application enters examination, it is routed to an Art Unit, a group of eight to fifteen patent examiners who review applications in the same technological field. Upon arrival, the Supervisory Patent Examiner (SPE) of that Art Unit randomly assigns the application to a specific examiner (Lemley & Sampat, 2012).[10] The examination will typically begin with the examiner conducting a prior art search, that is a search of previous patents, patent applications, or other publications, that are material to the patentability of the relevant invention. Upon completion of such search, the examiner assesses the patentability of the invention based on the criteria outlined in the Patent Act—e.g., whether the claimed invention is novel and nonobvious.

After assessing the patentability of the claims, an examiner composes a "first office action on the merits" (FOAM), which is non-final in nature (if it constitutes a rejection), and either allows the patent to issue or outlines the reasons for why the invention fails to meet the patentability standards. An applicant responds to a FOAM rejection by amending the claims or disputing the

[10] Occasionally, SPEs make non-random assignments, but in those instances, they do so not based on any characteristic that would affect the patentability of the application but instead, for instance, on an examiner's backlog of applications We conducted a series of telephone interviews with former SPEs to confirm these details of patent examination assignment.

rejection. Upon receipt of the applicant's response, the examiner will issue a second office action that will either: (1) allow the patent to issue, (2) finally reject the application, or (3) non-finally reject the application. If the examiner set forth all of the reasons for rejection in the FOAM and believes the invention still fails to meet the patentability standards, she will finally reject the application. If the examiner issued an incomplete FOAM—i.e., failed to include all bases for rejecting the application in the FOAM—she will issue a second non-final rejection which includes new grounds to reject the application.[11] The Patent Office views second non-final office actions as an indicator of low quality examination because an examiner who issues such an office action is essentially conceding her initial review was inadequate. Whereas an applicant can respond to a second non-final rejection by amending the claims or arguing the rejection is improper, an applicant's response is more circumscribed with respect to a final rejection. In the latter case, the aggrieved applicant must abandon the application altogether, appeal the denied application to Patent Trial and Appeal Board, or continue the examination process by filing a repeat application.[12]

Although the stages associated with the patent examination procedure are relatively structured, it is well recognized that patent examiners are afforded substantial discretion on how they approach the process (Cockburn et al. 2003). As with any complex task associated with substantial discretion, concerns regarding worker procrastination may arise. Allegations of procrastination within the Patent Office have recently made headlines. However, even before these latest controversies, the Agency was possibly aware of potential time-inconsistent behavior of examiners given their choice to utilize a conventional method to minimize procrastination—quotas.

[11] An examiner may not render a final rejection that contains a new ground of rejection, unless the new ground is necessitated by the applicant's claim amendments, or, in certain cases, if the new ground is based on information submitted in the applicant's information disclosure statement. MPEP 706.07(A).

[12] Repeat applications generally fall in one of two categories: continuation applications and Requests for Continued Examination (RCE). While there are technical differences between the two, which in part account for the popularity of the latter over the former, they are largely used for the same purpose: providing the applicant who has been denied the coverage she seeks with an additional chance for her patent application to be allowed.

III.B. Quotas

Examiners are expected to attain a certain number of work credits, often referred to as "counts," on a bi-weekly and quarterly period, where such expectations are a function of the complexity of the field in which the examiner is working and on her position in the general schedule (GS) pay scale.[13] Credits, however, have historically been earned only upon the issuance of a FOAM and at final disposal, which occurs when a patent application is either allowed by the examiner or abandoned by the applicant (often after receipt of a final rejection or in anticipation of such a rejection).[14] Notably, no work credits are earned for the issuance of a second non-final rejection.

Supervisors monitor examiners progress towards meeting their quotas at both bi-weekly and quarterly periods. In fact, examiners at pay grades GS-13 and below must have their decisions reviewed by a supervisor before they are communicated with the applicant. Examiners who have reached pay grade GS-13 with partial signatory authority have the ability to sign off on FOAM independently whereas examiners at pay grades at GS-14 and above have the ability to sign off on all of their work independently. In order to be promoted, an examiner typically need not only meet her workload goals but surpass them. Failure to earn the target amount of work credits can ultimately result in termination of employment (or in production bonus implications).

III.C. Hypotheses

III.C.1. Procrastination patterns

[13] A patent examiner in a more complex field has a lower quota of work units. Moreover, the higher the pay grade of an examiner within a technology area the greater her workload goals. Examiners must also meet workflow or docket management goals which seek to ensure that the flow of patent applications through the examination process align with prescribed time periods set by the Patent Office. These workflow goals are described in more detail in the Online Appendix. Because the workflow goals overwhelmingly align with end of bi-week, we utilize the term bi-weekly quota in the Article to refer to both production and workflow targets.

[14] Since 2010 examiners can also earn partial credits for final office actions and examiner-initiated interviews with the patent applicant or her attorney. Under either system, a patent examiner earns a maximum of two credits per patent application examined.

Given the role of quotas in personnel outcomes at the Patent Office, we assume that examiners will be incentivized to hit their production targets.[15] The question facing us is how they space out their work efforts over the observation period to reach this target. In the Online Appendix, we set forth a model inspired by Fischer (2001), which predicts that a time-consistent examiner that is motivated as such will reach her goals while roughly smoothing her work efforts evenly throughout the observation period (a prediction that is intuitive in light of the assumed concavity in utility for leisure). We extend this framework in the Online Appendix to introduce quasi-hyperbolic discounting by examiners (Laibson 1997). The key implication of this time-inconsistent approach is that examiners will tend to delay their initial intentions to begin working towards their bi-weekly goal and cluster their work efforts near the deadline.[16] In the first exercise of our empirical analysis below, we begin to test for the presence of procrastination by assessing whether examiners indeed bunch their work product around the end of the bi-weekly (and quarterly) quotas.

III.C.2. Consequences of procrastination

We offer two competing theories for the manner in which examiner procrastination may compromise the quality of application review. Each theory starts with an assumption that procrastinating examiners will face time constraints in completing the large number of tasks that they have left for themselves at the end of the quota period. First, examiners who are scrambling to meet their workload goals may grant patents excessively. That is, because patent applications are presumed valid—and thus no justification for an acceptance is required—an examiner who does not have sufficient time to conduct her search of the prior art and analyze the patentability of the claims may grant an invalid patent that she may not have otherwise granted.

[15] This assumption is supported by our interviews with examiners and former Supervisory Patent Examiners.
[16] This implication holds whether examiners are treated as naive or sophisticated hyperbolic discounters.

Alternatively, an examiner who has procrastinated may choose to meet her quota by issuing low quality or "shotgun" rejections at the end of a bi-week or quarter, which may likewise take less time to process than an otherwise unaffected review. A "shotgun" rejection is an invalid rejection that fails to meet the legal requirements—e.g., rejecting an invention as non-novel even though the cited prior art does not disclose the invention.[17] Notably, examiners may opt for this approach over the summary-acceptance approach when completing a FOAM given that examiners afford themselves a second chance in such instances to more thoroughly review the application in subsequent round(s). Simply accepting the application outright affords them no such opportunity.

IV. ANALYSIS

IV.A. Preliminary Analysis: End-loading of Applications

IV.A.1. Methodology and Data

Observationally documenting procrastination in the workplace is an exercise that confronts a number of methodological challenges. First necessary for such purposes are data on the work product of employees over sufficiently fine-grained intervals of time—e.g., daily data on employee output, where this output is amenable to quantification in the first place. Information of this nature allows one to depict a time path of employee behavior. To meet this need, we collected data from the Patent Office's Patent Application Information Retrieval (PAIR) database on nearly 2 million utility patent applications that were filed between March, 2001 and July 2012. The PAIR database provides information on a number of characteristics of, and events associated with, each application. Critical for our purposes is the PAIR's Transaction History File, which, among many

[17] A shotgun rejection is a term of art in the field of patent law. The term refers to patent examiners rejecting claims for "questionable reasons" in part "because of time pressures of work at the [Agency]" (Pressman and Stim 2015). In addition to providing broad-based empirical support regarding the existence of "shotgun rejections," our analysis also contributes to the field's understanding of this term of art by theorizing a more nuanced mechanism behind the time-pressured explanation for this behavior.

other things, provides information on the timing of completion—to the day—of the FOAM, along with information on the disposition of that FOAM—i.e., rejection or allowance.

Daily recordings of the completion of worker tasks alone, however, will generally not suffice to inform on the likelihood of procrastinating tendencies. Of critical import to a procrastination analysis is some benchmark signifying the time path that a non-procrastinating worker would follow in approaching her tasks. Naturally, establishing this benchmark requires some expectation as to how long the relevant task should take to complete given no delays in work effort. For instance, if one observes a journalist taking 14 days to write a story, it is difficult with this information alone to know whether the journalist worked diligently throughout that 14-day period or whether she did nothing for the first 13 days. Unlike this journalism hypothetical, our analysis of the patent context benefits from the fact that the Patent Office sets expectations—and enforces such expectations—as to how much work product examiners must complete over a 14-day period and thus how much time on average examiners must work on their reviews. While an enforced quota system is thus helpful in establishing this benchmark, one may be concerned that it removes the scope for procrastination in the first place. However, the fact that the Patent Office does not enforce expectations daily, but only over bi-weeks, leaves open the possibility that examiners will delay their work efforts within this evaluation period. The model that we present in the Online Appendix suggests that a non-procrastinating examiner would tend to smooth her work efforts within this bi-week period, whereas a procrastinating examiner would bunch her work efforts near the end of the period. Accordingly, the methodological thrust of this initial exercise is to assess whether examiners indeed bunch—i.e., end-load—their examination reviews in this manner.

IV.A.2. End-loading Results

Figure I presents a histogram of the completion of FOAM, broken down into daily frequency bins.[18] For initial illustrative purposes, we have chosen to do this over a representative year—2010. The daily frequency distribution depicted in Figure I demonstrates a striking degree of bunching in the completion of FOAM at the end of quota periods. Specifically, this figure evidences 26 evenly spaced spikes in the frequency by which FOAM are completed, coinciding with the end of each bi-week period. Examiner reviews are lightest at the beginning of each bi-week period. As the bi-weekly quota period nears its end, workload counts gradually begin to increase, with a spike on the last day of the bi-week period, a progression in daily productivity counts that is consistent with delayed onset of work efforts.[19]

In Figure II, we present a more generalized histogram that depicts FOAM counts by day, but where the relevant time period signifies days prior to a bi-weekly quota period event. Accordingly, this figure presents FOAM counts for 14 days, effectively averaging FOAM counts for all calendar dates that fell 13 days prior to a bi-weekly deadline, all calendar dates that fell 12 days prior to a bi-weekly deadline and so on and so forth. In this process, we include data from the full 2002-2012 sample. Figure II paints essentially the same picture as that shown in our representative year, 2010, from Figure I (note that the dips in activity represent weekends; the final day of the bi-weekly period generally falls on a Monday).

As stated previously, in addition to bi-weekly monitoring, examiners are expected to hit quarterly targets, effectively allowing them to catch up on any missed bi-weekly targets. As demonstrated by Figure I, we also observe a second degree of FOAM bunching that appears right

[18] Importantly, by FOAM, we refer to the first office action associated with a given application. Technically, the first office action after the filing of a Request for Continuing Examination (RCE) may be viewed as a FOAM. Since that action would be continuing the same application, however, we view it as different from the actual initial office action.
[19] The OIG reports (2014, 2015) previously mentioned do not depict histograms of this nature. Rather, they simply indicate that specified percentages of examiners submit specified percentages of their applications at the end of the quarter. As such, the OIG reports also overlook the bi-week aspect of workload expectations.

at quarter ends. Moreover, we observe a progression of increased intensity of bi-week spikes as we approach the end of the quarter—a progression that mirrors the daily patterns we observe within each bi-week. These patterns collectively suggest that examiners may delay in the onset of their biweekly targets in addition to delaying in the catch-up process they are afforded on a quarterly basis. Moreover, the magnitude of this end-of-period bunching is substantial. Nearly half of all FOAM completed in the dataset took place on the last day of a quota period (Table I).

IV.B. From End-loading to Procrastination: Overview of Next Steps

While end-loading of applications is consistent with the predictions of a model of examiner behavior characterized by time inconsistent preferences—and thus suggestive of procrastination—end-loading itself could conceivably be explained by a range of alternative time-consistent theories of behavior in which examiners work diligently and consistently throughout the sample period. For instance, among other theories we will consider, we may observe end-loading of FOAM reviews because patent examiners systematically work more hours on each application than expected by the Patent Office, causing them to rush at the end of the period on those that they did not leave themselves time for. The remainder of this paper attempts to set forth an additional span of empirical findings that collectively support a procrastination interpretation of examiner behavior and that mediate against various alternative time-consistent theories of behavior. To this end, we proceed by assuming that examiners do indeed procrastinate on their examination tasks and then predicting a range of additional behavioral outcomes—beyond the mere end-loading of applications depicted in Figure I—that one would expect to observe in the face of such procrastination. We then use our application level data and employ various methodological techniques to test each such prediction. In sub-section E, we then return to a discussion of a range of alternative time-consistent theories and address the ways in which the various findings favor a procrastination interpretation of the observed behaviors over these alternative theories.

IV.C. End-loading across Stages of Examination Process

Again, Figure I focuses on the first substantive review that patent examiners undertake. Given that only 11% of applications are allowed on the FOAM, the typical application involves multiple stages of review. To the extent that there are fixed costs associated with reviewing a file—e.g., an investment of time into understanding the basics of the claimed invention—one might predict that the burden of the examination process to the examiner diminishes across office actions. That is, the examiner may view the FOAM—where these fixed costs will be concentrated—as more unpleasant than the second and third office actions (should matters proceed that far). O'Donoghue and Rabin (2008) model a task process that involves multiple stages, much like the patent examination process (along with a model characterized by hyperbolic discounting). They suggest that one would predict more procrastination in the earlier stage of the work process if more of the fixed costs associated with the task are concentrated on that earlier stage.

These considerations might suggest more procrastination at the FOAM stage relative to later office actions. That is, we predict that examiners would tend to space their efforts on later office actions more evenly throughout the bi-week observation period. We assess this prediction in Figures A2 and A3 of the Online Appendix, replicating the histogram presented in Figure I but focusing on the second and third office actions, respectively (conditional on those applications that reach those stages of review). We continue to demonstrate substantial end-loading at the second-office-action stage, though nonetheless to a weaker extent relative to the first round of review. While nearly 50% of the FOAMs are completed at the quota end, 35% of the second office actions are completed at those times. Moreover, we find that the second dimension of end-loading—that is, clustering of reviews at the end of quarters—appears to have diminished considerably by the second stage of the review process. By the time applications proceed to the third stage of review (which is generally the first office action following the filing of a Request for Continued Examination), examiners tend to end-load their work efforts to a substantially weaker extent—only 11 percent of the time (note—lack of true end-loading would nonetheless entail 10 percent of reviews falling on the last day). Ultimately, this declining severity of end-loading by office action is consistent with the predictions just set forth and thus provides further evidence supportive of a procrastination interpretation for the end-loading of applications observed in Figure I.

IV.D. Evaluating the Consequences of End-loading / Procrastination

IV.D.1. Methodology

If examiners do procrastinate in their examination practices—continuing to assume that examiners are nonetheless motivated to hit their quota—one would predict that examiners would be forced to rush their remaining reviews at the end of the quota period. In Section II, we theorized various ways in which this end-of-period rushing may impact the quality of the examination process. In this sub-section, we test for markers of these theorized quality outcomes. To the extent that our evaluations of the outcomes of the examination process coincide with the predicted outcomes of a procrastination-induced end-of-period rush, the analysis may further lend support to a procrastination interpretation of the previously documented end-loading of applications. Beyond helping to mediate between time-consistent and time-inconsistent explanations for our findings, this analysis also offers novel insights generally regarding the consequences of workplace procrastination and specifically regarding the consequences to patent policy of examination delays.

To recap, we predict that procrastination-induced rushes at the end of the period may leave examiners inclined to issue quickly produced and ill-conceived "shotgun" rejections, leaving themselves the option to correct this inadequate review in subsequent rounds of review. To explore this possibility, we estimate the following specification out of the sample of first office actions:

$$FOAM\_Allow_{ait} = \alpha + \boldsymbol{\gamma}_{\mathbf{i}} + \boldsymbol{\lambda}_{\mathbf{t}} + \beta_1 Endloaded_{ait} + \boldsymbol{\beta}_2\mathbf{X}_{ait} + \varepsilon_{ait} \quad (1)$$

where *a* indexes the individual application, *i* indexes the individual examiner, *t* indexes the year in which the first office action on the merits (FOAM) is completed by the examiner. $FOAM\_Allow_{ait}$ indicates whether or not the application was allowed on the given FOAM. $Endloaded_{ait}$ indicates whether or not the FOAM was completed on the last day of the quota period. Year fixed effects (based on the timing of the FOAM) are captured by $\boldsymbol{\lambda}_{\mathbf{t}}$ and examiner fixed effects are captured by $\boldsymbol{\gamma}_{\mathbf{i}}$. With the latter inclusion, we effectively estimate whether given examiners allow applications on the FOAM at lower rates when those applications fall during the end-of-period time crunch relative to when they review applications during the middle of the quota period. Examiner fixed effects help alleviate concerns that examiners with high propensities to end-load fundamentally differ in their practices—e.g., in their allowance propensities—relative to examiners with smaller propensities to end-load. We also control for a number of other characteristics of applications, $X_{\mathbf{ait}}$, several of which the literature has demonstrated are important determinants of the granting practices of examiners (Frakes & Wasserman, 2017a): dummy variables for (1) the examiner's General-Schedule pay level, (2) examiner experience levels (in 2-year bins), (3) technology groups (using the 37 National Bureau of Economic Research technology sub-categories, (4) the incidence of a large entity applicant and (5) the incidence of foreign priority for the given application (previous filing at the JPO or EPO).[20] In sub-section F(3) below, we consider an instrumental variables modification to this specification to account for other unobservables.

[20] Standard errors are clustered to correct for autocorrelation within given examiners over time.

IV.D.2. Results

We present the results of this exercise in Table II. These findings are consistent with the "shotgun" rejection theory whereby examiners become less likely to allow on the FOAM when they are reviewing a large cluster of applications during the end of the quota period. On average, a given examiner will allow on the FOAM at a roughly 10 percentage-point lower rate—or a nearly 87 percent lower rate—when that same examiner reviews an application on the last day of a quota period relative to the prior days within the quota period. The magnitude of this relationship is staggering, reflecting a potentially substantial behavioral response to the end-of-period rush.

One might predict that any effect of end-of-period time crunches on the FOAM decision for an individual application will be especially strong in the case of those examiners facing substantial end-of-period demands on their full workload—i.e., those examiners most prone to end-loading practices. To test this prediction, we modify equation to include an interaction between the FOAM end-loading indicator for the given application and a variable capturing the relevant examiner's overall end-loading rate (calculated leaving out the given application). We present the results from this specification in Table II. Through the estimated coefficient of the interaction term, we test the prediction that the negative relationship between end-loading and FOAM allowances will only be stronger in the face of examiners with high mean end-loading rates—i.e., we predict a negative estimate of the coefficient of the interaction term. As demonstrated by Table II, we find evidence consistent with this prediction, providing greater confidence that the observation of substantially lower FOAM allowance rates for applications reviewed at quota ends derives from the hypothesized mechanism—i.e., an end-of-period time crunch.

To further evaluate whether this substantial elevation in end-of-quota rejections are indeed quick "shotgun" rejections whereby examiners buy time to conduct a more thorough review later, we proceed to evaluate the nature of the examination process in subsequent periods. To begin, we turn to the decisions made in the second round of review. Particularly, we look for markers indicative of the type of inadequate review in the first round that one might expect if examiners did indeed issue shotgun rejections in response to procrastination-induced time crunches. As noted above, the Patent Office views second non-final rejections as an indicator of low quality review because an examiner that issues such an office action is essentially conceding that her initial review of the application was incomplete. As demonstrated by Table III, we find that an application that was rejected in an end-loaded FOAM is around 1.4-2 percentage points—or roughly 8-12%—more likely to receive a second non-final office action rejection relative to an application whose initial rejection occurred during the within-quota period. This result is consistent with a theory that examiners are issuing low quality rejections at the end of quota period.

We next examine the likelihood that an application that is rejected on an end-loaded FOAM will ultimately be allowed throughout the course of the entire examination process. To the extent that procrastinating examiners are conducting more substantive reviews in later rounds one might predict that the ultimate allowance rates of an application with an end-loaded FOAM would approach that of an application whose FOAM was not end-loaded. Though end-loading appears to be associated with a substantially lower rate of allowance on the FOAM, there indeed appears to be a much weaker relationship between end-loading of applications and the ultimate likelihood that the application is allowed upon final disposition. We demonstrate such findings in Columns 4-6 of Table III, estimating specifications identical to (1) above but replacing the FOAM allowance incidence as the dependent variable with the incidence of the application ultimately being allowed (out of the sample of applications disposed of during the sample period). Applications whose FOAM were reviewed during an end-loaded period are roughly 1.8 percentages points—or roughly

2.5 percent—less likely to be allowed than those whose FOAM were reviewed during the within-quota period. These differences are far less than the gaps in allowance patterns present on the FOAM themselves (presented above).

This pattern of results is consistent with the theory that end-of-quota-period time constraints may induce examiners on the FOAM to submit weak and easily overcome rejections, affording themselves the ability to do a proper review on later iterations of the review period and thus the ability to exercise roughly the same degree of application scrutiny they otherwise would if they had not procrastinated in the first place (below, we offer an explanation for the small negative relationship that we nonetheless do find between end-loading and ultimate allowance).

In another attempt to assess the scrutiny of review applied to applications initially end-loaded, we look beyond the mere allowance or not of the application. Instead, we consider a metric indicative of the legal validity of any issued patent resulting from that application. For these purposes, we exploit the fact that many U.S. applicants likewise file for patent protection in European Patent Office (EPO) and the Japan Patent Office (JPO), two foreign offices that have roughly similar patentability requirements but invest substantially more in the examination process per application relative to the U.S. Patent Office. Accordingly, we consider the sample of issued patents in which the relevant U.S. applicant likewise sought protection at the EPO and JPO and use outcomes at these foreign offices as a benchmark—although an imperfect one—to assess the underlying validity of those patents issued by the U.S. Patent Office (Frakes and Wasserman, 2017a; Lemley and Sampat, 2012). If the ultimate review of applications whose FOAM were end-loaded were of the same degree of quality relative to those that were not end-loaded, we would expect the likelihood a U.S. issued patent was allowed at both the EPO and JPO would not depend upon whether the FOAM was end-loaded in the U.S. We present results from this exercise in

Columns 7-9 of Table III, finding a 0.6 percentage-point reduction—or roughly a 1.3 percent reduction—in the likelihood of a U.S.-issued patent being allowed at both the EPO and the JPO when that U.S.-issued patent was end-loaded on its FOAM. This suggests that those applications subjected to an end-of-quota rush may be of slightly weaker quality / legal validity relative to those applications not subjected to such rushes. Though end-loading may be associated with some quality deterioration, the magnitude of this difference is not substantial, consistent with the above-stated suggestion that the initially vast gap in the scrutiny of review between end-loaded and non-end-loaded FOAM narrows as examiners proceed to later stages of the review process.

Though procrastination does not appear to lead to substantial impacts on the way in which legal patentability standards are ultimately applied to an application, this behavioral phenomenon may nonetheless prolong the review process, an outcome with undesirable effects of its own. Put simply, if procrastination-induced time pressures are causing examiners to produce "shotgun" rejections on the FOAM, the result may be a completely wasted first-round of review. Understanding the extent to which examiner procrastination is contributing to application processing delays is of critical import to the Agency, given that the Patent Office has repeatedly stated that its growing backlog of applications is its biggest challenge.

To explore this mater, we estimate specifications that examine the relationship between the end-loading of a FOAM and various metrics indicative of the length of the review process: (1) the incidence of a Request for Continued Examination filed in connection with the application, a device that allows rejected applicants to continue their applications even after receiving a final rejection, (2) the number of Requests for Continued Examinations filed, and (3) the duration in days of the examination period. In each case, we find evidence suggesting that end-loading on the FOAM is associated with a prolonging of the examination review process (see Table IV). On average, an application that is reviewed at the end of the quota period relative to the within-quota period is roughly 2.4 percentage points more likely—or nearly 10 percent more likely—to have an RCE filed. We also find that end-loading of the FOAM is associated with a 12 percent increase in the number of RCEs filed for the application in question. Finally, end-loading on the FOAM is associated with an average increase of 50 days in the examination period.[21] In Section IV below, we calibrate the magnitude of these findings to suggest the degree to which examiner procrastination is contributing to the annual growth rate in the Agency's backlog.

IV.E. Telecommuting Analysis

The Patent Office has recently made a substantial push towards allowing patent examiners to work from home. We predict that examiner self-control problems will intensify upon entering a work environment where monitoring and supervision is decreased. To the extent it supports a self-control story in general, a finding that end-loading (and other markers of procrastination) increases upon the commencement of teleworking will further bolster the suggestion that the end-of-period clustering of FOAMs depicted in Figure I is itself the result of self-control limitations.

The Patent Office began rolling out its teleworking program in the mid-2000s. While the Agency offers a variety of telework options for its employees, the largest of these programs is the Patents Hoteling Program (PHP). To be eligible for the PHP, patent examiners must have achieved a GS-12 level, have positive performance ratings, and have worked at the Agency for at least two years. Participating patent examiners work from home at least 4 days a week and relinquish their office space at the Agency's headquarters. The PHP began in 2006 with 500 examiners and has been popular. By 2011 the program expanded to include over 2,600 examiners. The election into

[21] These duration findings perhaps explain the small, negative relationship that we do observe between the end-loading of a first office action and the ultimate allowance of that application. The longer the duration of the examination, the more likely it is that an applicant will abandon her application. Since abandonments are considered an application disposition, this possibility may be contributing to the observation of a slightly lower overall grant rate for end-loaded versus non-end-loaded applications.

the hoteling program does not affect an examiner's quota. For each examiner in our dataset, we obtained information—via the filing of Freedom of Information Act Requests—on whether or not they participated in the PHP and what day, month and year they started working from home.

To test the above prediction, we estimate the following examiner fixed-effects specification:

$$Endloading_{ait} = \alpha + \boldsymbol{\gamma}_{\mathbf{i}} + \boldsymbol{\lambda}_{\mathbf{t}} + \boldsymbol{\beta}_1 \sum_{r=-4}^{4} \mathbf{Telecommuting}_{ir} + \boldsymbol{\beta}_2 \mathbf{X}_{ait} + \varepsilon_{ait} \qquad (2)$$

where *a* indexes the individual application, *i* indexes the individual examiner, *t* indexes the year in which the first office action is completed by the examiner, and where $\mathrm{Endloading}_{ait}$, $\boldsymbol{\lambda}_{\mathbf{t}}$, $\boldsymbol{\gamma}_{\mathbf{i}}$. $\mathrm{X}_{\mathbf{ait}}$, are as above. We include a series of event-time dummy variables ($\sum_{r=-4}^{4} \mathbf{Telecommuting}_{ir}$) indicating the periods of time leading up to and following a given examiner's onset of telecommuting—that is, a dummy variable indicating that an application's first office action was performed in the 4th year prior to the onset of telecommuting, the 3rd year prior to the onset of telecommuting, and so on and so forth. This specification essentially embraces a dynamic difference-in-difference framework, in which we look at changes in end-loading practices before and after an examiner's commencement of telecommuting, using those examiners not switching their teleworking status over those time periods as a control group. The dynamic aspect of this specification allows us to explore whether any such change in end-loading tendencies precipitated the start of the telecommuting program—which would undermine a causal interpretation of the findings—and whether any response in behavior to telecommuting evolved over time.

We plot the coefficients of the series of event-time indicators for the telecommuting variable in Figure III (in Table A1 of the Online Appendix, we show the corresponding regression table). As this figure demonstrates, in the period of time leading up to the commencement of telecommuting, the telecommuting and non-telecommuting examiners trended in the same direction. Immediately upon joining the telework program, however, an examiner's incidence of end-loading—of completing a given application's first office action on the last day of the quota period—increases by 7 percentage points, or by nearly 14%.[22] In Table A3 of the Online Appendix, we demonstrate the year-by-year stability in the incidence of two immutable application characteristics—applicant entity size and foreign priority status—as examiners approach and surpass the onset of telecommuting. This falsification exercise lends confidence to our telecommuting research design in suggesting no case mix sensitivity to an examiner's telecommuting experiences.

Further supporting the notion that this reduced supervision intensifies examiners' procrastinating tendencies, the evidence also suggests that the degree to which examiners issue "shotgun" rejections at the end of the quota increases following the onset of telecommuting. Previously, we flagged the presence of shotgun rejections by looking within given examiners and comparing first office action allowance rates for those applications that they review at the end of the quota period relative to those that they review during the interim quota period. In Table V, we extend this shotgun-rejection exercise by tracking how such comparisons evolve in the years leading up to and subsequent to the onset of telecommuting, which we accomplish by estimating examiner fixed effects specifications that include each of the event time indicators, an indicator variable for the application falling at the quota end, and the interaction between the event time indicators and the end-of-period indicator. The estimated coefficients of these interaction terms suggest that the FOAM-allowance differential between the end of period and the interim period intensifies following telecommuting—that is, even more rejecting on the FOAM for end-loaded relative to

[22] We note that this pattern of coefficients looks nearly identical when taking a more balanced approach that only follows examiners that we can follow for at least 4 years before and after telecommuting.

non-end-loaded applications—despite the fact that the corresponding differential in the odds that the application is ultimately allowed does not substantially change over this time.[23]

All told, these findings suggest that the pattern of results presented in the above sections only strengthens when examiners face less direct supervision. This finding reinforces a procrastination interpretation of the above patterns in light of the expected effect that the weakened supervisions associated with telecommuting would tend to have on pre-existing self-control problems.

While the shock to the examiners' end-loading practices upon the onset of telecommuting seems clear from the findings, there is some uncertainty in the data regarding how long-lived this effect is. First, we observe that the difference in end-loading rates between the treated and control examiners returns to its pre-telecommuting levels after several years, suggesting that examiners may learn some degree of discipline in working from home over time. Interestingly, though, the rate by which end-loaded applications are rejected on the FOAM relative to non-end-loaded applications continues to remain high—perhaps even increases further—several years following telecommuting. As such, telecommuting may actually lead to a more permanent shift in the tendency to rush work efforts at the end of the quota period.

We acknowledge the possibility that worker preferences themselves may change upon shifting to a home-working environment. When facing requirements to spend a set number of hours per day at a centralized office, workers' opportunities for leisure on those days are arguably limited—e.g., web-surfing, reading, gossiping with co-workers, etc. In such environments, the assumption of diminishing returns to leisure made in the above model is perhaps reasonable. With diminished supervision and the ability to work out of a home environment, workers acquire the ability to enjoy

[23] The downward trend in the sequence of interaction coefficients arguably begins prior to the onset of telecommuting; however, the negative trend in the FOAM allowance differential between end-loaded and non-end-loaded applications intensifies following the beginning of telecommuting, consistent with a causal response.

new forms of leisure during the work week, including acts of leisure that may require an investment of a critical block of time—e.g., amusement parks, skiing, etc. With this new possibility set, workers may begin to exhibit some degree of convexity in preferences for leisure and therefore desire to cluster work efforts on certain days in order to free up a sufficient degree of time on other days. Accordingly, from a time-consistent framework, one might predict some degree of increased clustering in work effort following telecommuting. Under this alternative framework, however, there would be no reason to predict that this new clustering would always occur on the last day of the Patent Office's quota period. The fact that deadline spikes only intensify following telecommuting favors the procrastination story over this convex-preferences alternative.

IV.F. Assessing Alternative Time-Consistent Explanations

Collectively, the above findings are consistent with the various predictions we have made about patent examiner behavior under a theory of procrastination. With this final analysis, we consider the possibility that these collective findings may also be consistent with a range of time-consistent models of examiner behavior. Each of the alternatives considered are arguably consistent with the mere fact that we observe clustering of applications at the end of quota periods. We assess, however, whether these alternative theories are consistent with the remaining empirical observations that we have made in our efforts to test the predictions of the procrastination model.

IV.F.1. Paper Pile Alternative

End-of-quota clustering in FOAM reviews could be consistent with a story in which examiners work diligently throughout the observation period but simply accumulate their work product until the point at which they turn in all of their reports at one time. Much of the additional findings, however, are inconsistent with any such story. As an initial matter, we note that this explanation is incomplete in light of the fact that nearly half of all reviews continue to be processed in the within-quota period. Second, consider our observation of even more intense spikes on quarter-ends. If examiners worked consistently over time and simply filed all of their reports at once, they would do so and process all reports at every moment in which they were obligated to do so—i.e., every two weeks—leaving no room for an independent quarter effect. Finally, there would be no reason under this alternative story to expect that (1) end-loading would intensify upon the onset of telecommuting, (2) that end-loaded applications would tend to exhibit markers suggestive of shotgun rejections, or (3) that end-loading would weaken with successive office actions.

IV.F.2. Supervisor Effect

One may also be concerned that the end-of-period spikes shown in Figure I are solely the result of delays in the supervisory approval process. That is, junior examiners may be working diligently throughout the observation period but their supervisors are only approving their work in bunches at the end of the quota period. As an initial matter, we note that this supervisor-delay alternative does not entail a rush in the examiner's work effort at the end of the period. As such, under this alternative view, one would not expect to observe that end-loaded applications exhibit markers indicative of shotgun rejections. The presence of such markers arguably favors the procrastination explanation. In any event, to more completely assess this concern, we separately estimate the degree of end-loading for examiners with and without the authority the sign off on their own FOAM. As demonstrated by Table A2 in the Online Appendix, even those examiners with independent authority of this nature exhibit a substantial degree of end-of-period bunching of application reviews, doing so nearly 44% of the time (compared with 49% for the entire sample), suggesting that the patterns from Figure I cannot be solely explained by a supervisor effect.

IV.F.3. Application Sorting

One may have concerns that the end-of-period bunching demonstrated in Figure I is exclusively the result of examiners sorting applications in various ways—for instance, examiners working on more complex and thus more time-consuming applications during the beginning of the bi-week, and then turning to the less time-intensive applications near the end.

To begin, we note that the observed increase in end-loading upon the onset of telecommuting arguably favors the procrastination story. The Patent Office's telecommuting program would neither be expected to increase the share of easy cases that are docketed to an examiner, nor increase the fundamental desire to sort based on complexity. Furthermore, the unique institutional nature of the patent examination process also mediates against this concern. The completion of the examination itself is tied to the tolling of a specified period of time—if the examiner is unable to find and articulate a basis for rejection in that specified time, they are expected to allow the application. As such, in the case of a difficult application that entails a cumbersome search into its novelty and non-obviousness, the assigned examiner is expected to end the relevant search after the designated time. This expectation might tend to cut against the possibility that the light FOAM counts early in the bi-weeks reflects examiners focusing on especially complex applications and spending more time on those applications than expected. Nonetheless, these expectations are only enforced on average via the bi-week quota mechanisms, in which case we acknowledge that examiners may nonetheless deviate from a per-application time expectancy. It bears emphasizing, however, that for any such sorting to produce the degree of end-loading that we observe, there would need to be an arguably unusual distribution of application complexities—a distribution in which nearly half of all applications are easy enough to address over a day's time, whereas the remaining half are complex enough to be spread over the remaining nine working days.[24]

It would also be unusual for this sorting story to produce the various markers of shotgun rejections that we have observed. That is, under this sorting alternative, the cluster of examinations at the end of the quota period would represent easier, less time-consuming cases. It is unclear why those easily reviewed applications would be characterized by substantially higher rates of rejection on the first office action followed by a leveling out of allowance rates in subsequent rounds of review. If the end-loaded applications were inherently easier (either easy rejections or easy allowances), one might expect that whatever allowance differential we observe on the FOAM would persist when viewing allowance rates on the application as a whole.

In any event, let us assume that there is indeed an unobservable feature of application—e.g., complexity—on which examiners are sorting and that might explain the pattern of results presented in Tables II and III. In a final empirical exercise, we attempt to account for any such unobservables by modifying the examiner fixed-effects specification from equation (1) to instrument the end-loading indicator variable for the given application with the examiner's overall end-loading rate for the given year (leaving out the influence of the given application on the end-loading rate calculation). To the extent the mean end-loading rate reflects an examiner's time-varying end-loading "style," one might expect that this style metric would correlate highly with the likelihood that an individual application is end-loaded. Moreover, in light of the fact that applications are randomly assigned to examiners, one would not expect that an examiner's mean end-loading rate to be associated with the unobserved complexity of the application at hand. As demonstrated by

[24] Moreover, very long patent applications also often contain multiple inventions in which an examiner can issue a restriction requirement before substantive evaluation which in essence forces the patent applicant to break the application down into multiple applications for review.

Table VI, when taking this instrumental variables approach, the estimated relationship between an end-loaded application and each of the following metrics indicative of the shotgun-rejection story is very similar to the estimates previously presented in Tables II and III: the incidence of allowance on the FOAM, the likelihood of a second office action non-final rejection, the likelihood of ultimate allowance of the application, the duration in days of the examination process.

Finally, we note that there are no meaningful differences between end-loaded applications and non-end-loaded applications in the two immutable application characteristics that we have in our data—i.e., the incidence of a "large-entity" applicant and the incidence of a prior application at either the European Office or the Japan Patent Office—as demonstrated by Table IV.

IV.F.4. Examiner Systematically Spending More Time Than Expected on Each Application

We consider the possibility that examiners will systematically deviate from the Patent Office's expectations by spending more time on each application than they are explicitly instructed to. If examiners begin each quota period in this manner, they will naturally leave themselves an insufficient amount of time at the end of the quota period to satisfy their internal desire to spend an above-expected amount of time on each application. In this scenario, it is possible that examiners may likewise decide to issue a shotgun rejection at the end of the quota period in order to buy themselves enough time during a subsequent round of review to put in the desired amount of effort. As such, this alternative explanation could account for the observation of both end-of-period bunching of FOAM reviews and the various markers of shotgun rejections. To be clear, should this alternative mechanism in fact account for the observed results, examiners would likely be spending double the amount of time on applications than instructed. Regardless of whether this hypothesized expenditure of time is plausible, we do note that this alternative story is not consistent with one important component to our empirical analysis—the telecommuting results.

IV.F.5. Clustered Work Styles

Lastly, it might be that when examiners are more confident of their analysis—i.e., the application appears to be a clear allowance or a clear rejection—they file the FOAM during the interim period but when the patentability of an application is less clear they hold onto the FOAM in hopes that they will be able to revisit the application before the bi-week is completed. In such a scenario, the end-of-period bunching we observe in Figure I could be the result of examiners preserving the option to reconsider a set of close call applications if time allows. This explanation seems incomplete as examiners would have to be on average holding approximately fifty percent of all FOAM in the hopes of returning to the applications later in the quota period. It seems unlikely that examiners would reserve the option to revisit that many applications considering that they would not have time to act on many of those options. Moreover, this alternative exaplanation is likewise one that does not predict a change in endloading upon telecommuting.

V. Conclusion

The practice of end-loading by patent examiners—or waiting until deadlines to submit required reviews—has been the subject of much recent conversation within the patent sphere and has even been the topic of evaluation in recent reports by the Commerce Department's Office of Inspector General (OIG 2014, 2015). In this paper, we have attempted to document the full extent of this practice and to investigate its origins. Does endloading arise from examiners procrastinating on their assigned tasks or from some time-consistent pattern of behavior? The OIG reports suspect that it is the former, but offers no analysis to help separate these possibilities Hypothesizing that examiners indeed procrastinate on their work tasks, we theorize various ways in which this behavior may impact their work product (beyond mere endloading itself). We also theorize that examiner procrastination will increase following a shift towarads examiner telecommuting. We find evidence consistent with each such prediction, supporting the procrastination hypothesis.

To be sure, we can not say with confidence that the entirety of the observed end-of-period bunching is a result of procrastination. It is possible that at least some of the spikes at the end of the quota are a result of a combination of the various time-cosistent alternative hypotheses that we explore. At the same time, however, the results suggest that the entirety of the observed end-loading cannot be accounted for by these alternative stories. Collectively, the results suggest that procrastination may be playing a strong role in shaping examiner behavior. If one nonetheless assumes that the full extent of the observed end-loading reflects procrastination behavior, equation (4) of the Online Appendix implies a present bias parameter of nearly 50%. This is comparable to the 30-40% present bias estimated in Bisin and Hyndman (2014) in their experimental analysis of procrastination among college students in performing designated tasks and to the 10-50% present bias estimated in a range of additional studies surveyed by DellaVigna (2009).

Our analysis further suggests that the most notable consequence of procrastination to the quality of the examination process is additional delay in the review process. Essentially, an inadequate review completed in a rushed state at the end of a quota period is a lost opportunity that must be made up for on subsequent rounds of review. Examination review delays may interfere with the deployment of valuable inventions to the marketplace and increase the uncertainty surrounding the rights of potential patents, which in turn may limit a company's ability to license or engage in related activity (Frakes and Wasserman, 2016). Delays may also lead firms to increasingly rely on other methods of appropriation such as trade secrets, thus preventing public disclosure of information that future inventors may build upon.[25] The literature has yet to exhaustively quantify the extent of any of these social welfare harms. Nonetheless, those harms may be extensive enough to have spurred the Patent Office into paying particular attention to its backlog. In fact, the Agency has repeatedly stated that its biggest challenge to fulfilling its mission of providing high quality timely review of applications is its existing backlog of applications (Frakes and Wasserman 2015).

Given that roughly half of all first office actions are end-loaded and that the Agency presently reviews approximately 500,000 applications a year, we estimate that examiner procrastination in just the first office action—let alone procrastination in subsequent rounds of review—is attributing to over 12 million additional processing days a year. Through a simple back-of-the-envelope calculation applied to this amount, our analysis implies that the increases in processing delays stemming from procrastination of first office actions may have contributed to as much as 17% of the highly publicized annual growth rate in the backlog of applications awaiting first review over our sample period.[26] As such, the welfare implications of this practice are potentially considerable.

By implementing externally imposed deadlines in the first place, the Patent Office has arguably already taken steps to confront the negative consequences that might arise from employee procrastination (Arielly and Wertenbroch 2002). That being said, our analysis cannot conclude whether the quota system that the Patent Office has chosen to implement is suboptimal. It is possible that the harms associated with examiner procrastination are outweighed by the costs associated with enforcing work efforts on a more granular level. Such costs could include

[25] While we do not find evidence that examiner procrastination directly leads to the issuance of patents that fail to meet the patentability standards, our previous work found evidence that the Patent Office may grant patents of questionable legal validity in an effort to decrease its application backlog (Frakes and Wasserman 2015). Thus, to the extent that examiner procrastination increases delays in processing applications then it may also—albeit indirectly—lead to the issuance of low quality patents and implicate the host of social welfare harms associated with allowing invalid patents to issue.

[26] To arrive at this number, we note that the 12 million in aggregate additional days in review may account for the processing of roughly 11,000 fewer applications per year in light of the mean number of days of the prosecution period for the individual applications in our sample (1,128). With the backlog of applications awaiting final review growing by nearly 66,000 per year over our sample period, this implies that the practice of end-loading on the FOAM may be contributing to roughly 1/6 of the annual growth rate in the Patent Office's backlog.

additional administrative time and effort, loss of flexibility to examiners to efficiently arrange their work activity, and possible decrease in overall job satisfication. Though hard to quantify, these latter costs may be meaningful in light of the difficulty the Patent Office may face in in attracting quality examiners when subjecting examiners to less discretion and more rigid workload goals.

Of course, the Patent Office may have softer, less expensive tools at its disposal than simply increasing the frequency by which it sets quota expectations, tools which may enable examiners to retain some of the flexibility of the current system. The Agency, for instance, could attempt to smooth work by incorporating a measurement of end-loading into the examiner's performance apprasial. Alternatively, if the real concern regarding examiner procrastination is its effect on the quality of patent examination, the Agency could adopt a system where a percentage of end-loaded applications are subject to further scrutiny by the Patent Office.

Relatedly, while our analysis also suggests that the practice of procrastination and the harms that ensue from it may intensify following the onset of telecommuting, it may nonetheless be true that the cost savings from telecommuting outweigh these harms. The Patent Office touts its teleworking program as "worth its weight in gold" stating that the program provided over $64.7 million in efficiency benefits to the Patent Office in 2014 alone.[27] These are substantial savings that should naturally be balanced against any costs stemming from increased examiner procrastination following the transition to telecommuting.

DUKE UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH
UNIVERSITY OF TEXAS

[27] U.S. Patent and Trademark Office, Telework Annual Report 2015.

REFERENCES

Acland, Dan, and Matthew R. Levy, "Naivete, projection bias, and habit formation in gym attendance," *Management Science* 61(1) (2015), 146-60.

Akerlof, George, "Procrastination and Obedience," *American Economic Review* 81(2) (1991), 1-19.

Ariely, Dan, and Klaus Wertenbroch, "Procrastination, Deadlines and Performance: Self-Control by Precommitment," *Psychological Science* 13 (2002), 219-24.

Asch, Beth J., "Do Incentives Matter? The Case of Navy Recruiters," *Industrial and Labor Relations Review* 43 (1990), 89S-106S.

Balasubramanian, Natarajan, Jeongsik Lee and Jagadeesh Sivadasan, "Deadlines, Work Flows, Task Sorting, and Work Quality," *Management Science* (2017, forthcoming).

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying, "Does Working from Home Work? Evidence from a Chinese Experiment," *Quarterly Journal of Economics* 130(1) (2015), 165-218.

Bisin, Alberto, and Kyle Hyndman, "Present-Bias, Procrastination and Deadlines in a Field Experiment," NBER Working Paper No. 19874 (2014).

Brown, Alexander, Zhikang Eric Chua and Colin .F. Camerer, "Learning and Visceral Temptation in Dynamic Saving Experiments," *Quarterly Journal of Economics* 124(1) (2009), 197-231.

Cadena, Ximena, Antoinette Schoar, Alexandra Cristea, and Heber M. Delgado-Medrano, "Fighting Procrastination in the Workplace: An Experiment," NBER Working Paper No. 16944 (2011).

Chetty, Raj, Emmanuel Saez, and Laszlo Sandor, "What Policies Increase Prosocial Behavior? An Experiment with Referees at the *Journal of Public Economics*," *Journal of Economic Perspectives* 28(3) (2014), 169-88.

Cockburn, I., S. Korum, and S. Stern, "Are All Patent Examiners Equal? Examiners, Patent Characteristics, and Litigation Outcomes," in *Patents in Knowledge-Based Economy*, W. M. Cohen and S.A. Merril (Eds.), (Washington, DC: National Academies Press, 2003).

Courty, Pascal and Gerald Marshke, "An Empirical Investigation of Gaming Responses to Performance Incentives," *Journal of Labor Economics*, 22 (2004), 23-56.

DellaVigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature* 47(2) (2009), 315-72.

Duflo, Esther, Michael Kremer, and Jonathan Robinson, "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," *American Economic Review* 101(6) (2011), 2350-90.

Fischer, Carolyn, "Read This Paper Later: Procrastination with Time–Consistent Preferences," *Journal of Economic Behavior and Organization* 46(3) (2001), 249–269.

Frakes, Michael D. and Melissa F. Wasserman, "Does the U.S. Patent and Trademark Office Grant Too Many Bad Patents?: Evidence from a Quasi-Experiment," *Stanford Law Review*, 67(3) (2015).

---------, "Patent Office Cohorts," *Duke Law Journal* 65 (2016), 1601-1655.

---------, “Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents?: Evidence from Micro-Level Application Data,” *Review of Economics and Statistics*, Forthcoming (2017a).

---------, “Empirical Scholarship on the Prosecution Process at the U.S. Patent Office,” in *Research Handbook on the Economics of Intellectual Property Law (Volume II—Analytical Methods)*, forthcoming (eds. Peter S. Menell and David L. Schwartz) (2017b).

Frederick, Shane, George Loewenstein, and Ted O’Donoghue, “Time Discounting and Time Preference: A Critical Review,” *Journal of Economic Literature* 40 (2002), 351-401.

Gruber, Jonathan, and Botond Koszegi, “Is Addiction “Rational”? Theory and Evidence,” *Quarterly Journal of Economics* 116 (4) (2001), 1261-1303.

Kremer, Michael, Supreet Kaur, and Sendhil Mullainathan, “Self Control at Work.” *Journal of Political Economy* 123 (6) (2015), 1227 - 77.

Laibson, David, “Golden Eggs and Hyperbolic Discounting,” *Quarterly Journal of Economics* 112(2) (1997), 443-77.

Laibson, David, Andrea Repetto, and Jeremy Tobacman, “Estimating Discount Functions with Consumption Choices over the Lifecycle,” NBER Working Paper 13314 (2007).

Larkin, Ian, “The Cost of High-Powered Incentives: Employee Gaming in Enterprise Software Sales,” *Journal of Labor Economics*, 32 (2014), 199-227.

Lemley, Mark A. and Bhaven Sampat, “Examiner Characteristics and Patent Office Outcomes,” *Review of Economics and Statistics* 94(3) (2012), 817–827.

Loewenstein, George, and Jon Elster, *Choice Over Time*, Russel Sage, New York (1992).

Moser, Petra, “How to Patents Laws Influence Innovation? Evidence from Nineteenth Century World’s Fairs,” *American Economic Review* 95(4) 2005, 1214-36.

O'Donoghue, Ted, and Matthew Rabin, “Doing it Now or Later,” *American Economic Review* 89(1) (199), 103-124.

--------, “Procrastination on long-term projects,” *Journal on Economic Behavior and Organization* 66 (2008), 161-75.

Office of Inspector General Report, *Abuse of Telework Program at the USPTO*, Inspector General Referral No. PPC-CI-12-1196-H (2014).

--------, *USPTO Needs to Strengthen Patent Quality Assurance Practices*, OIG-15-026A (April 10, 2015).

Oyer, Paul, “Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality,” *Quarterly Journal of Economics*, 113 (1998), 149-185.

Pressman, David & Richard Stim, Nolo’s Patents for Beginners, 8th ed., Berkeley, CA, Nolo (2015)

Read, Daniel, George Loewenstein and Shobana Kalyanaraman, “Mixing virtue and vice; combining the immediacy effect and the diversification heuristic,” *Journal of Behavioral Decision Making* 12(4) (1999), 257-73.

Shapiro, Jesse M., “Is there a daily discount rate? Evidence from the food stamp nutrition cycle,” *Journal of Public Economics* 89 (2005), 303-25.

Sprenger, Charles, “Judging experimental evidence on dynamic inconsistency,” *American Economic Review: Papers & Proceedings* 105 (5) (2015), 280-85.

TABLE I

Summary Statistics

	(1)
	MEANS (STANDARD DEVIATION)
Incidence of Application End-loading (First Office Action on the Merits Completed on Last Day of Quota Period)	0.483 (0.499)
Incidence of First Office Action Allowance	0.114 (0.318)
Incidence of Non-Final Rejection on Second Office Action	0.165 (0.371)
Incidence of Application Allowance (After Final Disposition)	0.713 (0.452)
Incidence of Request for Continued Examination	0.232 (0.422)
Number of Requests for Continued Examination	0.263 (0.616)
Examination Duration in Days (Time between Filing and Final Disposition)	1129.263 (526.622)
Incidence of U.S. Patent being Allowed at both the JPO and EPO	0.446 (0.497)
Incidence of Large-Entity Applicant	0.727 (0.445)
Incidence of Application being Filed at the EPO or JPO prior to U.S. Filing	0.071 (0.258)
Incidence of Application being Reviewed while Examiner Telecommuting	0.113 (0.317)

Statistics are from a sample of 1,741,500 first office actions completed between March, 2001 and July, 2012 (focusing only on applications filed after March, 2001).

TABLE II

Relationship between End-loading of First Office Actions and First-Office-Action Allowance Rates

	(1)	(2)	(3)	(4)	(5)	(6)
End-loaded Application Dummy	-0.108*** (0.002)	-0.104*** (0.002)	-0.104*** (0.002)	-0.022*** (0.006)	0.004 (0.005)	0.002 (0.005)
End-loaded Application Dummy X Examiner-End-loading Rate	-	-	-	-0.162*** (0.013)	-0.215*** (0.010)	-0.212*** (0.011)
Examiner Fixed Effects?	NO	YES	YES	NO	YES	YES
Year Effects and Other Covariates?	NO	NO	YES	NO	NO	YES

Standard deviations are indicated in parenthesis and are clustered at the examiner level. Estimates are from a sample of 1,741,500 first office actions completed between March, 2001 and July, 2012 (focusing only on applications filed after March, 2001). Coefficient of examiner end-loading rate in Column 4 omitted.

TABLE III

Relationship between End-loading of First Office Actions and Various Future Outcomes of Application

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	INCIDENCE OF SECOND-OFFICE ACTION NON-FINAL REJECTION			INCIDENCE OF ULTIMATE ALLOWANCE OF APPLICATION			INCIDENCE OF APPLICATION BEING ALLOWED AT BOTH THE EPO AND JPO		
End-loaded Application Dummy	0.037*** (0.002)	0.019*** (0.001)	0.014*** (0.001)	-0.045*** (0.002)	-0.018*** (0.001)	-0.018*** (0.001)	-0.004* (0.002)	-0.006*** (0.002)	-0.006*** (0.002)
Examiner Fixed Effects?	NO	YES	YES	NO	YES	YES	NO	YES	YES
Year Effects and Other Covariates?	NO	NO	YES	NO	NO	YES	NO	NO	YES

Standard deviations are indicated in parenthesis and are clustered at the examiner level. Estimates are from a sample of 1,741,500 first office actions completed between March, 2001 and July, 2012 (focusing only on applications filed after March, 2001). Estimates in Columns 7-9 are from a sub-sample of this initial sample, focusing on those application that culminated in an allowance at the U.S. Patent Office and whose underlying inventions were also the subject of an application at both the European Patent Office and the Japan Patent Office.

TABLE IV

Relationship between End-loading of Applications and Various Other Outcomes and Measures

	(1)	(2)	(3)	(4)	(5)	(6)
	EXAMINATION DURATION ANALYSIS (LENGTH OF TOTAL EXAMINATION PERIOD)				FALSIFICATION EXERCISES (IMMUTABLE APPLICATION CHARACTERISTICS)	
	INCIDENCE OF REQUEST FOR CONTINUED EXAMINATION	LOG NUMBER OF REQUESTS FOR CONTINUED EXAMINATION	LOG DURATION OF EXAMINATION PERIOD	DURATION IN DAYS OF EXAMINATION PERIOD	ENTITY SIZE OF APPLICANT	FOREIGN PRIORITY STATUS OF APPLICANT
End-loaded Application Dummy	0.024*** (0.001)	0.120*** (0.005)	0.049*** (0.001)	49.394*** (1.336)	0.002** (0.001)	0.001** (0.000)

Standard deviations are indicated in parenthesis and are clustered at the examiner level. Estimates are from a sample of 1,741,500 first office actions completed between March, 2001 and July, 2012 (focusing only on applications filed after March, 2001). All regressions include examiner and year effects and various covariates.

TABLE V

Relationship between Examiner Telecommuting and the Differential in Various Measures between End-loaded and Non-End-loaded Applications

	(1)	(2)
	Allowance on FOAM	Ultimate Allowance of Application
End-loaded Application	-0.094***	-0.019***
	(0.003)	(0.001)
(Omitted: > 4 Years Prior to Telecommuting Dummy)		
4-Years Prior to Telecommuting Dummy	-0.008*	-0.012***
	(0.004)	(0.004)
3-Years Prior to Telecommuting Dummy	0.006	-0.006
	(0.005)	(0.004)
2-Years Prior to Telecommuting Dummy	0.017***	-0.012***
	(0.005)	(0.005)
1-Year Prior to Telecommuting Dummy	0.026***	-0.008
	(0.006)	(0.005)
Year Starting Telecommuting Dummy	0.060***	0.007
	(0.007)	(0.006)
1-Year Post Starting Telecommuting Dummy	0.074***	0.013**
	(0.008)	(0.006)
2-Years Post Starting Telecommuting Dummy	0.080***	0.017***
	(0.009)	(0.006)
3-Years Post Starting Telecommuting Dummy	0.092***	0.028***
	(0.010)	(0.007)
>= 4-Years Post Starting Telecommuting Dummy	0.080***	0.033***
	(0.012)	(0.008)
(Omitted: > 4 Years Prior to Telecommuting Dummy X End-loaded Application)		
4-Years Prior to Telecommuting Dummy X End-loaded Application	0.024***	0.010***
	(0.004)	(0.004)
3-Years Prior to Telecommuting Dummy X End-loaded	0.008*	0.002

Application	(0.005)	(0.003)
2-Years Prior to Telecommuting Dummy X End-loaded	-0.004	0.005
Application	(0.005)	(0.003)
1-Year Prior to Telecommuting Dummy X End-loaded	-0.023***	-0.002
Application	(0.005)	(0.003)
Year Starting Telecommuting Dummy X End-loaded	-0.064***	-0.011***
Application	(0.006)	(0.004)
1-Year Post Starting Telecommuting Dummy X End-loaded	-0.094***	-0.013***
Application	(0.007)	(0.004)
2-Years Post Starting Telecommuting Dummy X End-loaded	-0.122***	-0.014***
Application	(0.008)	(0.005)
3-Years Post Starting Telecommuting Dummy X End-loaded	-0.163***	-0.017***
Application	(0.010)	(0.006)
>= 4-Years Post Starting Telecommuting Dummy X	-0.178***	-0.027***
End-loaded Application	(0.013)	(0.007)

Standard deviations are indicated in parenthesis and are clustered at the examiner level. Estimates are from a sample of 1,741,500 first office actions completed between March, 2001 and July, 2012 (focusing only on applications filed after March, 2001). All regressions include examiner and year effects and various covariates.

TABLE VI

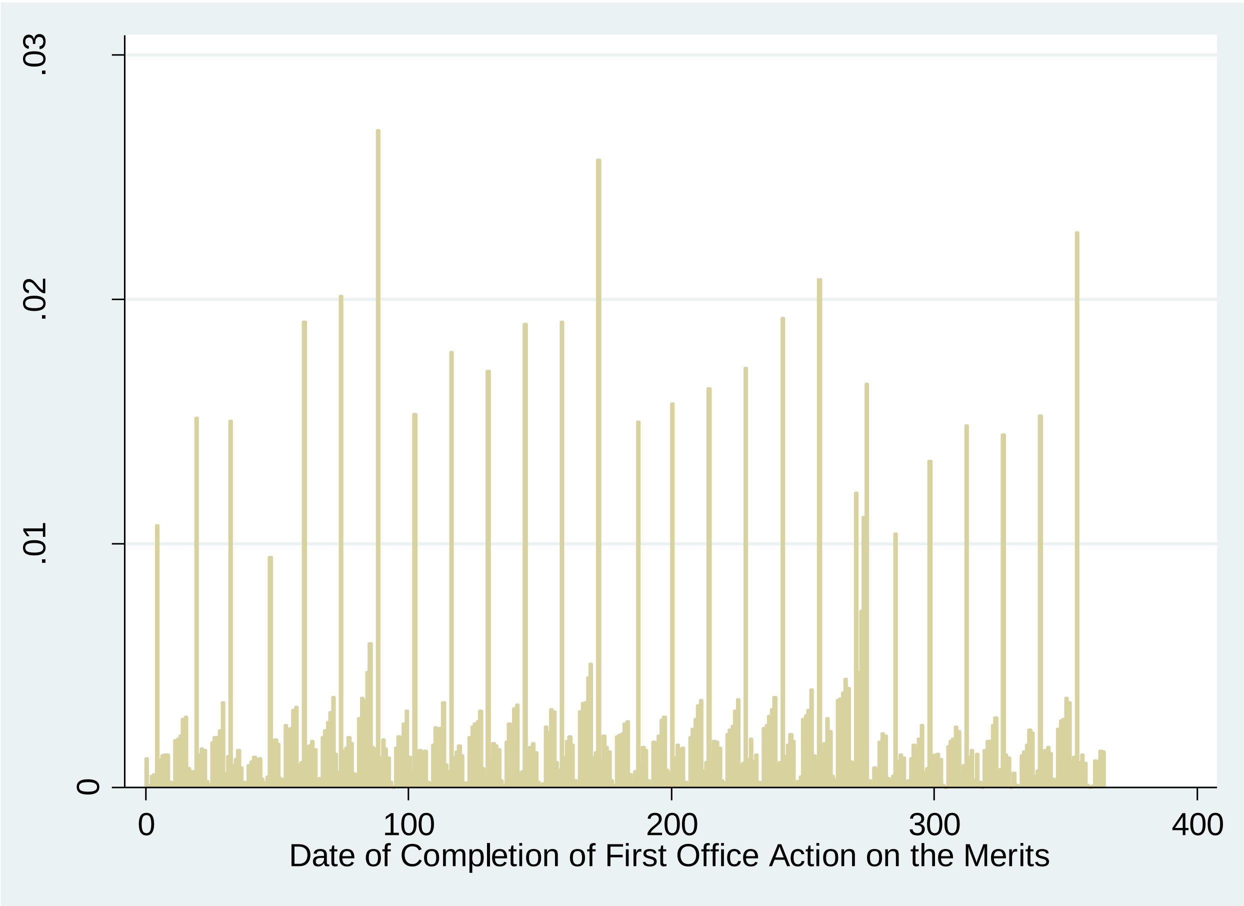
Relationship between End-loaded Applications and Various Application Outcomes: Instrumental Variables Estimates

	(1)	**(2)**	**(3)**	**(4)**
	FOAM Allowance	Ultimate Allowance of Application	Second-Office-Action Non-Final Rejection	Examination Duration (in Days)
End-loaded Application Dummy	-0.076***	-0.022***	0.010***	34.55***
	(0.002)	(0.003)	(0.002)	(1.99)

Standard deviations are indicated in parenthesis and are clustered at the examiner level. Estimates are from a sample of 1,741,500 first office actions completed between March, 2001 and July, 2012 (focusing only on applications filed after March, 2001). The indicator variable representing whether or not the application's first office action was reviewed on the last day of the quota period is instrumented by the examiner's mean end-loading rate among all applications that they reviewed during the given year (not considering the contribution of the application at hand). All regressions include examiner fixed effects, year fixed effects and various controls representing characteristics of the application and examiner.

FIGURE I

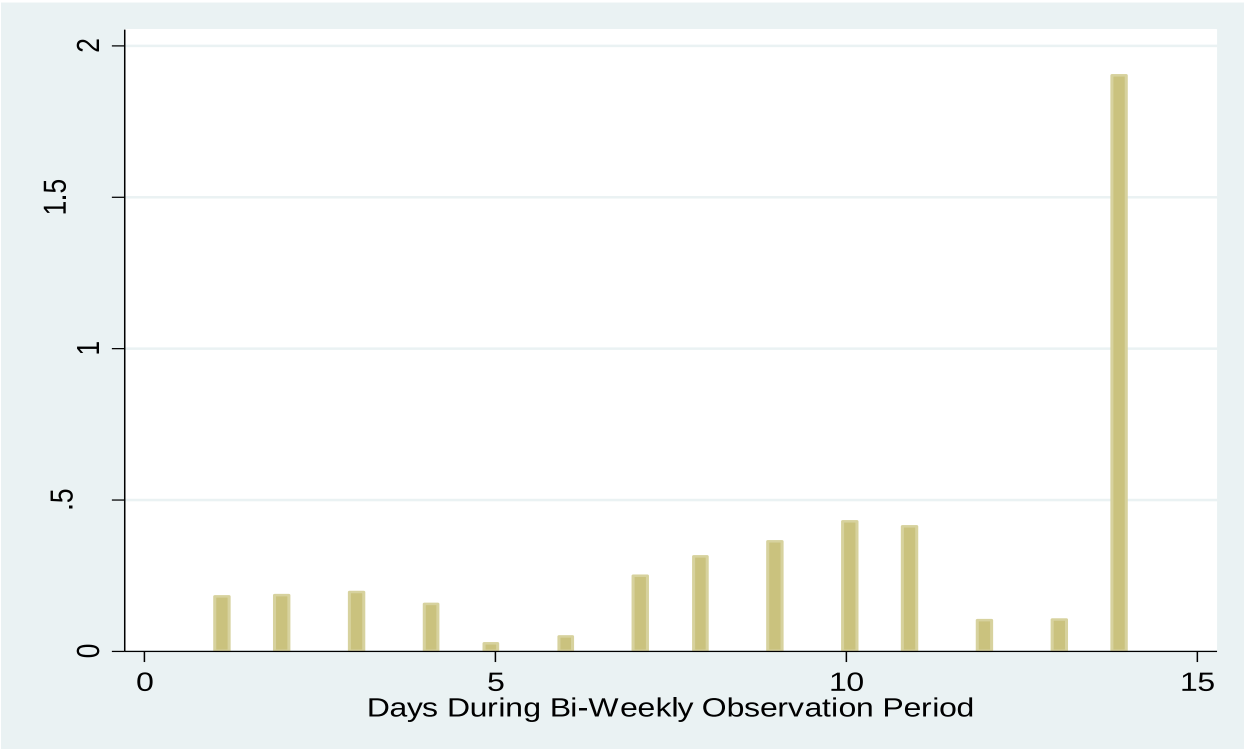
Frequency Distribution of First-Office-Action Reviews across Individual Days in 2010



Notes: Frequency counts are from the universe of FOAMs completed during the course of 2010 and were obtained from the Transaction History File of the Patent Office's PAIR database.

FIGURE II

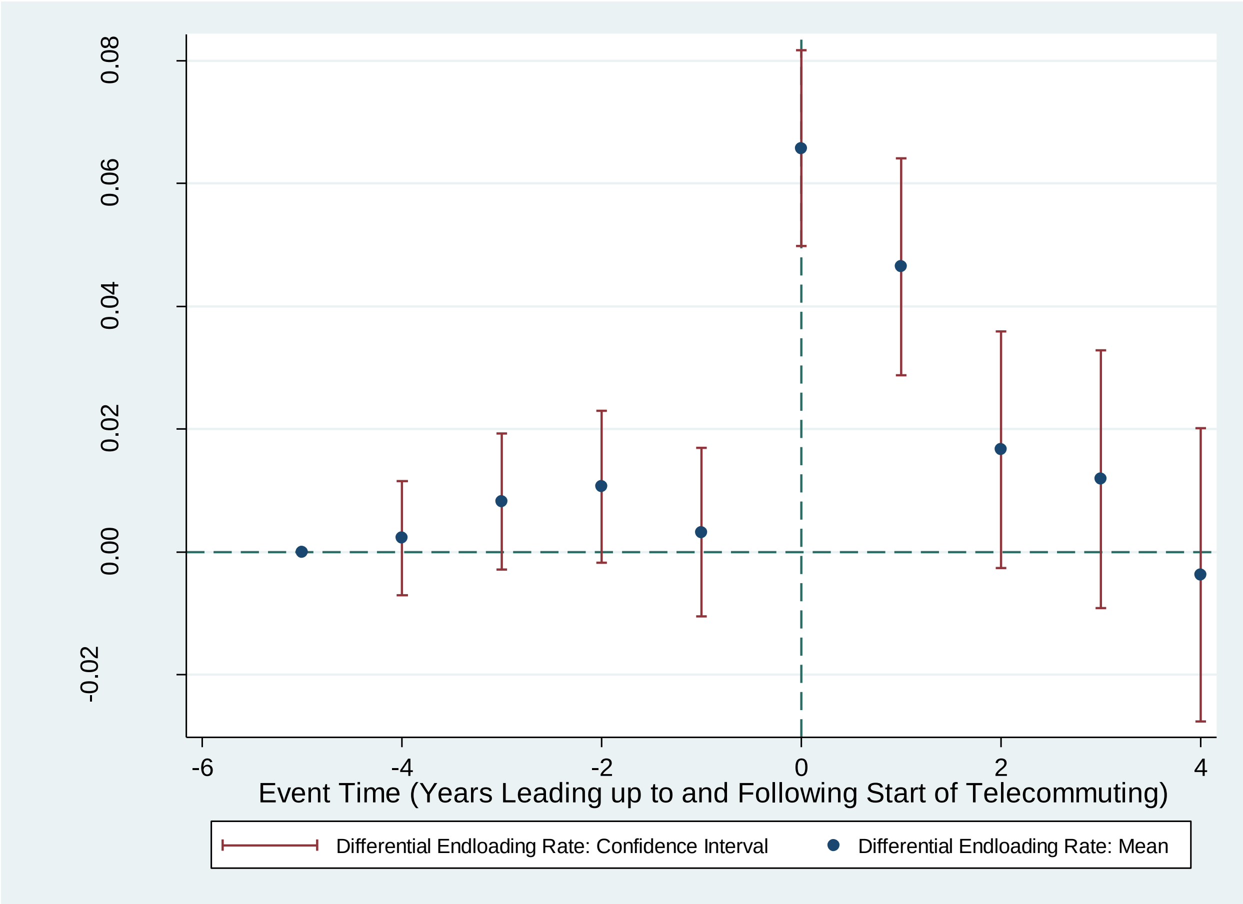
Frequency Distribution of First-Office-Action Reviews by Days Prior to End of Bi-Weekly Quotas



Notes: Frequency counts are from the universe of FOAMs completed during the course of 2001-2012 and were obtained from the Transaction History File of the Patent Office's PAIR database.

FIGURE III

Event Study Analysis: Relationship between Examiner Telecommuting and the Likelihood that First Office Actions are Completed on the Last Day of the Quota Period



Note: this figure presents coefficients from a regression of the incidence of an application's FOAM being end-loaded on a series of dummy variables representing leads and lags of the associated examiner's commencement of telecommuting (if at all). The regression includes year and examiner effects and covariates representing various characteristics of the applications. Standard errors are clustered at the examiner level. Estimates are from a sample of 1,741,500 first office actions completed between March, 2001 and July, 2012 (focusing only on applications filed after March, 2001).

ONLINE APPENDIX

PROCRASTINATION IN THE WORKPLACE: EVIDENCE FROM THE U.S. PATENT OFFICE

MICHAEL D. FRAKES AND MELISSA F. WASSERMAN

Details on Docket Management Process

In addition to meeting production quotas, patent examiners are also expected to meet workflow or docket management goals. The workflow or docket management goals seek to ensure that the flow of patent applications through the examination process align with prescribed time periods set by the Patent Office. Patent examiners have five different dockets of patent applications, wherein each docket contains patent examinations in a different stage of review. More specifically, these five dockets include: (1) new patent applications; (2) patent applications which have been amended in response to an office action; (3) patent application for which a final office action has been sent; (4) patent application in which a supervisor has issued a correction; (5) patent application on an accelerated examination schedule. Each docket has an "expected average days" for completion. Once a patent application is docketed, a clock begins to count down from the expected average days for review for that docket. Examiners are expected to complete review of an application before its workflow clock expires. As a result, examiners must meet both production quotas—complete a certain number of work credits every bi week—and workflow goals—complete stages of examination review in certain timeframes. Notably, workflow goals largely align with production quotas. That is, the expected average days for completion overwhelmingly expire at the end of a production bi-week. Thus, we refer to the quota in our paper to encompass both production and workflow goals.

Model of Patent Examiner Behavior

A) Time-consistent benchmark

As a starting point, given the role of quotas in personnel outcomes at the Patent Office, we assume that patent examiners will be incentivized to hit their bi-weekly (and quarterly) production targets.[28] The question facing us is how they space out their work efforts over the observation period to reach this target. One can readily predict that a time-consistent examiner that is motivated as such will reach her goals while roughly smoothing her work efforts evenly throughout the observation period. For these purposes, we set forth a model inspired by Fischer (2001), which is well suited for our needs in that it contemplates a situation where the execution of a task requires a number of hours to complete, where those hours can be spread out over a designated period of time. With this model, we attempt to predict the time path that a rational, time-consistent patent examiner will follow over the course of the 10 working days in a bi-week period.

On any given day, examiners receive utility of u(24-h) for receiving 24-h hours of leisure, where h represents the number of hours spent that day reviewing applications and where u is strictly increasing and concave. Examiners are expected to spend 80 hours over the bi-week reviewing applications, though are not obligated to spend 8 hours each day on such tasks. We assume however, that they are monitored enough that they spend greater than 0 hours per day on examinations (to avoid discussing the other corner solution possibility, we also assume that examiners will not spend the full 24 hours of any given day reviewing applications). B^t represents the exponential discount factor. Examiners select the amount of time spent reviewing applications on each of the 10 working days in the bi-week (h_0, h_1, …h_9) by solving the following:

[28] This assumption is supported by our interviews with examiners and former Supervisory Patent Examiner.

$$\underset{h_0, h_1, \dots h_9 \in (0, 24)}{Max} \quad U = \sum_{t=0}^{9} u(24 - h_i)\, \beta^t$$

s.t.

$$\sum_{t=0}^{9} h_t = 80$$

The first order conditions from this problem suggest the following relationship:[29]

$$u'(24\text{-}h_0) = u'(24\text{-}h_1)\beta = \dots = u'(24\text{-}h_9)\beta^9 \qquad (1)$$

For any $\beta < 1$, in light of the concavity of u, it is readily apparent from these first order conditions that the number of hours selected will increase to some extent over time as the 10^{th} day approaches. The following equation demonstrates the degree to which hours worked change over time:[30]

$$\frac{h_{t+1} - h_t}{h_t} \approx \frac{(1-\beta)}{\beta}\left(\frac{u'(24 - h_t)}{-u''(24 - h_t) * h_t}\right) \qquad (2)$$

The first term on the right simply captures the degree to which examiners prefer the present. If examiners do not discount at all over this short time period, this term equals zero and examiners do not change their hours day-by-day and instead smooth their work efforts over time. The second term captures the elasticity of intertemporal substitution for leisure (EIS). Essentially, whatever change in the temporal work path brought about by general time preferences is mediated by the degree to which examiners will entertain deviating from a smooth leisure profile over time.

Noting that the literature generally estimates an EIS of less than 1, Fischer (2001) conservatively assumes an EIS of 1 and thereafter suggests that a daily change in hours worked of just 1% would

---

[29] This follows naturally from the fact that each first order conditions suggests $u'(24\text{-}h_t)\beta^t = \lambda$, for each t.

[30] To derive this equation, we follow Fischer (2001) and start with the observation that $u''(24\text{-}h_t) \approx -\Delta u'(24\text{-}h_t) / \Delta h_t = -(u'(24\text{-}h_{t+1})\text{-}u'(24\text{-}h_t)) / (h_{t+1} - h_t)$. From this, we derive $(h_{t+1} - h_t) / h_t = --(u'(24\text{-}h_{t+1})\text{-}u'(24\text{-}h_t)) / u''(24\text{-}h_t)h_t$. The next step is to replace the numerator of the right-hand-side of this preceding equation. For these purposes, we note that the first order conditions from the above maximization problem suggests: $u'(24\text{-}h_t)\beta^t = u'(24\text{-}h_{t+1})\beta^{t+1} = \beta^{t+1} (u'(24\text{-}h_{t+1}) - u'(24\text{-}h_t)) + \beta^{t+1} (u'(24\text{-}h_t)$. Reorganizing, this suggests that $u'(24\text{-}h_{t+1}) - u'(24\text{-}h_t) = ((\beta^t - \beta^{t+1}) / \beta^{t+1}) * u'(24\text{-}h_t)$. Inserting this into the above equation, we find: $(h_{t+1} - h_t) / h_t = -((1 - \beta) / \beta) * (u'(24\text{-}h_t) / (u''(24\text{-}h_t)*h_t)) = ((1 - \beta) / \beta) * (u'(24\text{-}h_t) / (-u''(24\text{-}h_t)*h_t))$.

require an annual rate of time preference ($\frac{(1-\beta)}{\beta}$) of 3800% or an annual β of a staggeringly low 0.026.[31] If instead one assumes a perhaps more reasonable annual β of 0.75, this analysis would suggest a near 0% daily increase in hours worked—i.e., a smooth time path in work effort over relatively short periods of time.[32] Accordingly, we predict that a patent examiner that discounts future utility exponentially and that has time-consistent preferences will tend to smooth her work efforts near evenly over the bi-week observation period (a prediction that is intuitive in light of the assumed concavity in utility for leisure).

B) Time-Inconsistent Predictions

The above framework can be extended to introduce sources of time-inconsistency in behaviors. For instance, examiners might discount future leisure in a quasi-hyperbolic manner (Laibson 1997). That is, an examiner at time 0 may discount leisure at time 1 at Bδ, leisure at time 2 at $B^2\delta$, and so on and so forth; essentially, in this framework, the examiner wants to discount tomorrow's leisure by Bδ, even though the examiner today wants her future self to follow normal, exponential discounting at B^t thereafter. The time inconsistency in behavior arises because tomorrow's examiner—when tomorrow arrives—will likewise tend to assign that additional δ discount for all periods beyond that date. Modifying the above framework to incorporate a present bias of this nature, examiners at time t solve the following:

$$\underset{h_t, h_{t+1}, \ldots h_9 \in (0,24)}{Max} \; U = u(24 - h_t) + \sum_{i=1}^{9-t} u(24 - h_{t+i})\, \beta^i \delta$$

s.t.

[31] O'Donoghue and Rabin (2015) provide a similar discussion.

[32] While the discount rate implied by an assumed 1% change in daily work effort may not comport with our expectations of general exponential time preferences, consider the discount rates implied by the degree of work effort changes we actually observe. In our analysis below, we find that nearly half of the work effort is completed at the end of the bi-week period suggesting as much as a 10% daily change in hours worked, which, under the same assumptions, would suggest an annual β of essentially 0 (reflecting a near complete preference for the present). This is possible but unlikely in the face of a rational, exponential discounter.

$$\sum_{i=1}^{10-t} h_i + S_t = 80$$

where S_t represents the inherited stock of hours worked from hours worked decisions in the time periods prior to t ($h_1 + h_2 + \ldots h_{t-1}$). Assuming that examiners are naïve hyperbolic discounters that are not aware in time t of the fact that examiners at time t+1 will also attach the additional δ discount factor to periods t+2 and beyond, it is straightforward (based on the above analysis) to show that the first order conditions from this problem imply the following:

$$u'(24-h_0) = u'(24-h_1)\beta\delta = \ldots = u'(24-h_9)\beta^9\delta \qquad (3)$$

which suggests:

$$\frac{h_1 - h_0}{h_0} \approx \frac{(1-\beta\delta)}{\beta\delta}\left(\frac{u'(24-h_0)}{-u"(24-h_0) * h_0}\right) \qquad (4)$$

With the presence of δ, which often is thought to capture a substantial degree of discounting of tomorrow, the degree to which examiners would discount time 1 at time 0 may now be considerable enough that one would predict a notable increase in the hours worked between today and tomorrow (as distinct from the discussion above).[33] The remainder of equation (3) suggests that the planned rate of change in hours worked over time from period 2 onwards will follow the rule set forth in equation (2) above. Of course, when period 2 arrives, the examiner will again apply a present bias in her optimization problem at that time, suggesting an hours worked amount in period 2 less than what she plans to apply in period 3, and so on and so forth. The implication of this pattern is that examiners will end-load their work efforts at the deadline, assuming again that examiners are ultimately motivated to hit their 80 hours per bi-week requirement of work.

[33] For instance, if one assumes a δ of 0.75 (again assuming an intertemporal elasticity of substitution of 1), one would expect a roughly 1/3 increase in work effort between time 0 and time 1.

In a companion to her 2001 paper, Fischer (2001) extends this framework to allow for a sophisticated hyperbolic discounter who, at time t, decides on her current hours allocation knowing that her future self tomorrow will likewise incorporate a δ discount between tomorrow and the next day. Interestingly, Fischer's model predicts an even greater degree of procrastination to the extent that a worker today knows that her future selves will heavily prioritize current leisure, thereby encouraging her to work even less today in order to force her future selves to work more.

Accordingly, while time-consistent examiners will tend to spread their work out evenly over the observation period, examiners with present-biased preferences (or examiners who exhibit differentially discounting) will tend to delay their initial intentions to begin working towards their bi-weekly goal and cluster their work near the deadline. In the first exercise of our empirical analysis below, we test for the presence of procrastination by first assessing whether examiners indeed bunch their work product around the end of the bi-weekly (and quarterly) quotas.

<u>Additional references</u>:

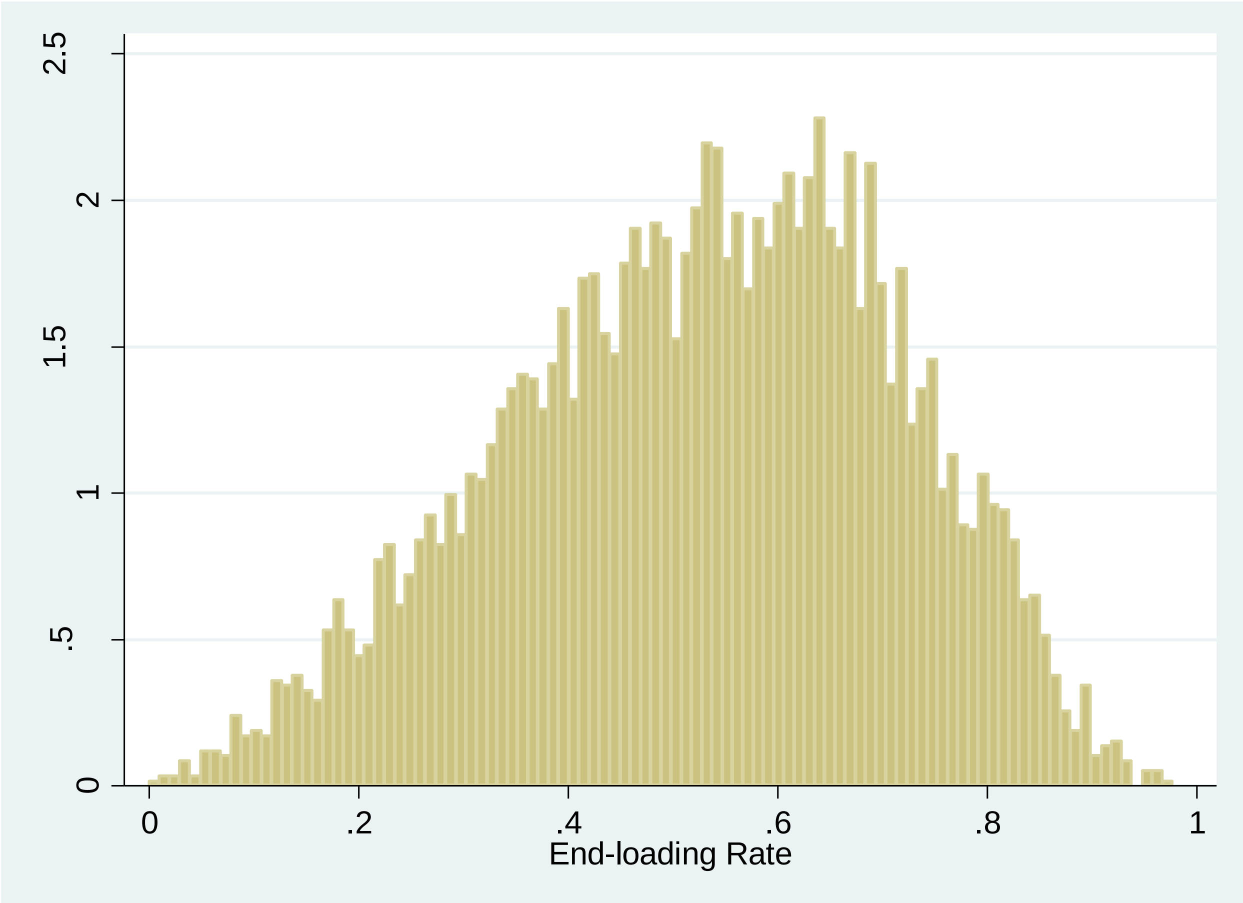
Fischer, Carolyn, "Read This Paper Even Later: Procrastination with Time–Inconsistent Preferences," RFF Discussion Paper 99–20 (2001).

End-loading Heterogeneity

To complement our demonstration of average end-loading in application reviews, we consider the possibility of heterogeneity in end-loading practices across examiners. To the extent this end-loading is indeed reflective of procrastination, we acknowledge that not all examiners are likely to exhibit the same degree of time inconsistency in work efforts. To assess the degree of heterogeneity in such behaviors, we calculate the mean rates by which each examiner completed a FOAM on the last day of a quota period and then present the distribution of these mean rates across examiners in Figure A1. Though there is considerable variance in end-loading rates across examiners, Figure A1 does demonstrate that the vast majority of examiners exhibit some striking degree of end-of-period clustering of reviews. Even at the 20th percentile of examiners (ranked according to their end-loading tendencies), nearly 35 percent of FOAMs reviewed fell on the last day of the quota period. If examiners were to smooth their workload over the 10 business days inherent in the bi-week period, one would instead expect to observe only 10 percent of applications being processed on the final day.

FIGURE A1

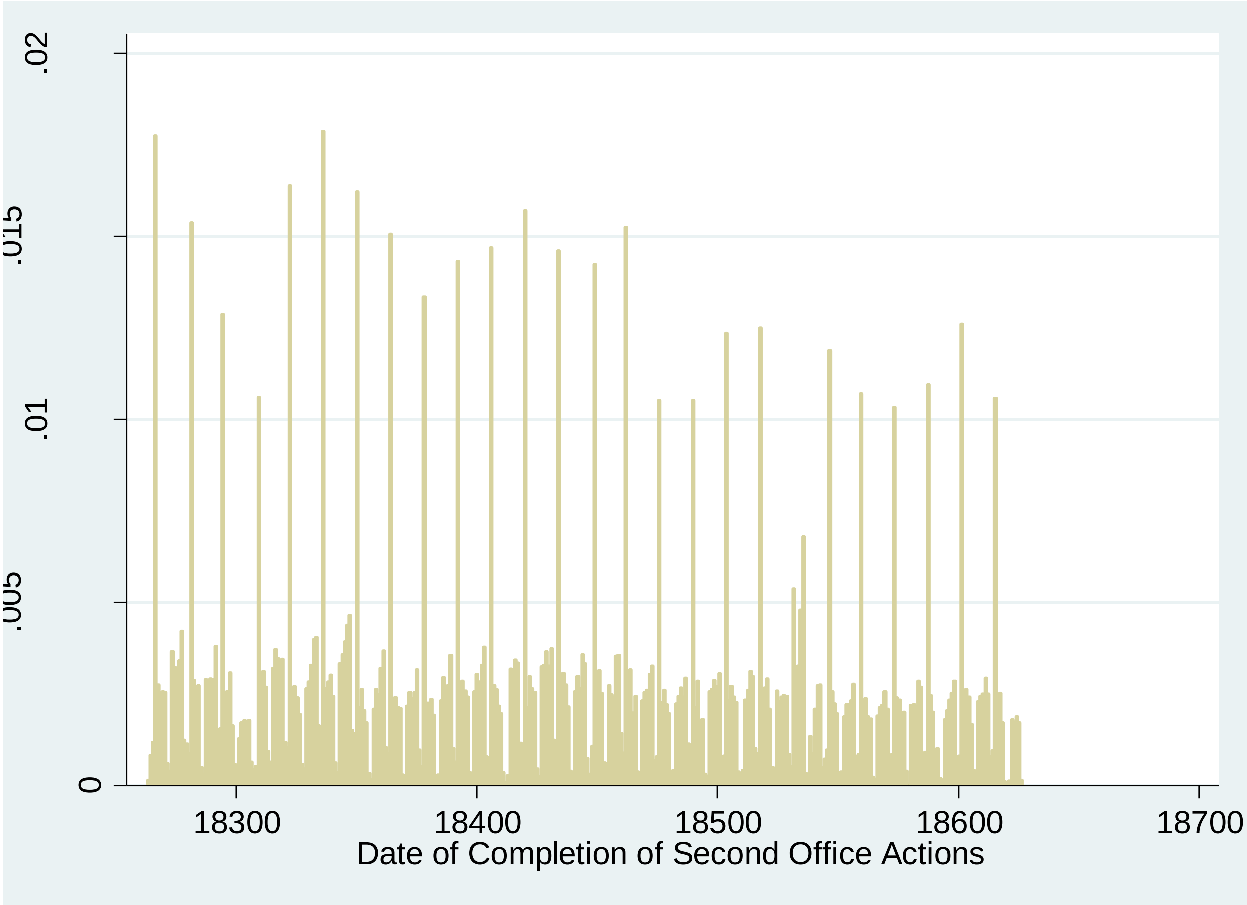
Distribution of Mean End-loading Rates across Examiners



Notes: this histogram is derived from a sample of 9,639 examiners completing first office actions represented in the Patent Office's PAIR database between March, 2001 and July, 2012 (for applications filed after March, 2001).

FIGURE A2

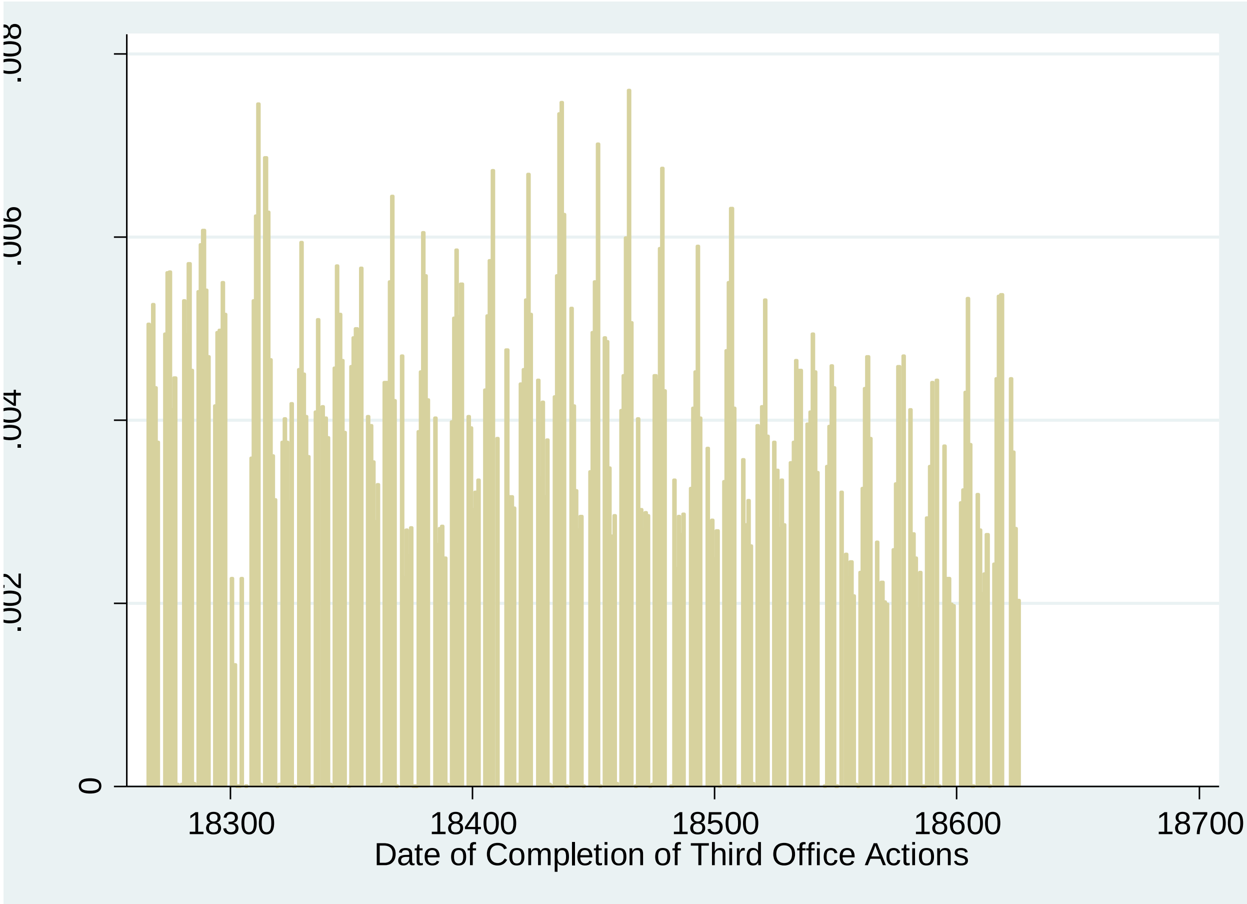
Frequency Distribution of Second-Office-Action Reviews across Individual Days in 2010



Notes: Frequency counts are from the universe of second office actions completed during the course of 2010 and were obtained from the Transaction History File of the Patent Office's PAIR database.

FIGURE A3

Frequency Distribution of Third-Office-Action Reviews across Individual Days in 2010

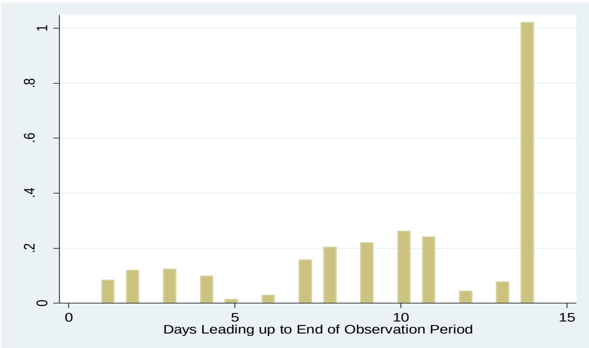


Notes: Frequency counts are from the universe of third office actions completed during the course of 2010 and were obtained from the Transaction History File of the Patent Office's PAIR database.

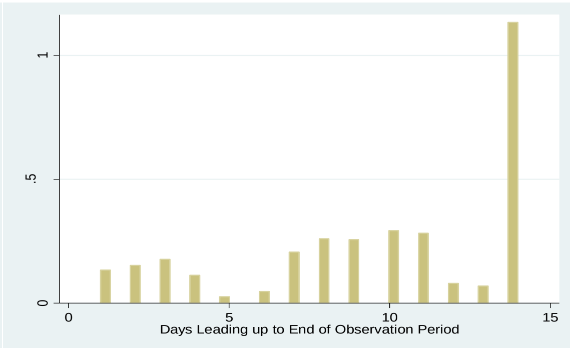
FIGURE A4

Generalized Bi-Weekly Histogram of FOAM Counts by NBER Technology Sub-Category

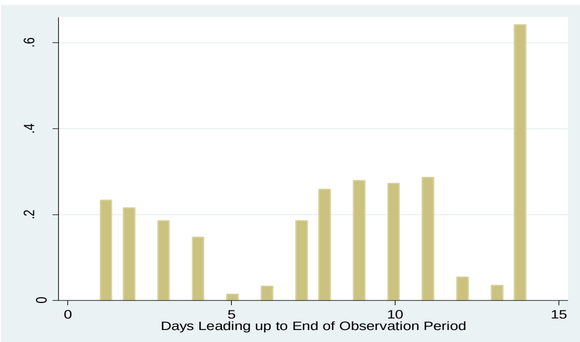
Agriculture Food and Textiles



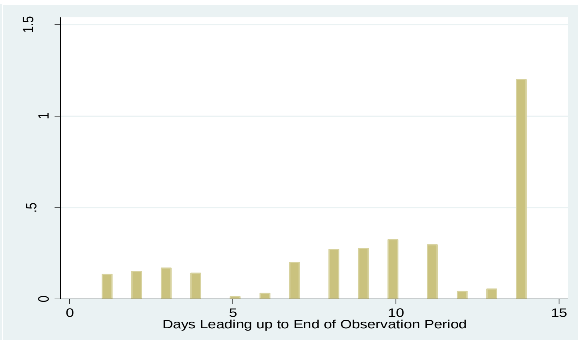
Coating



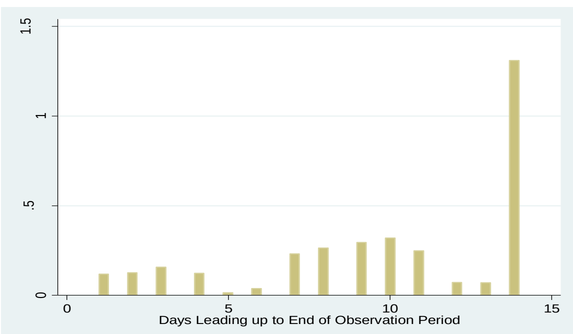
Gas



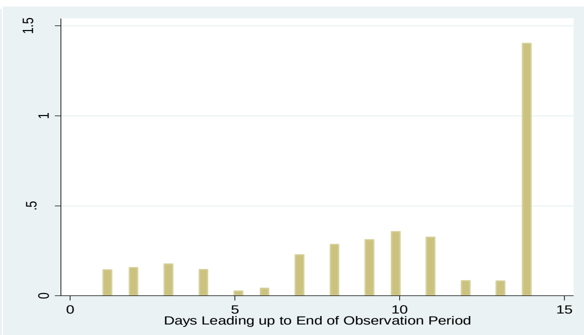
Organic Compounds



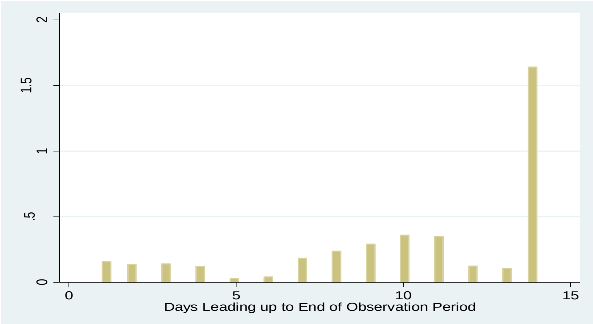
Resins



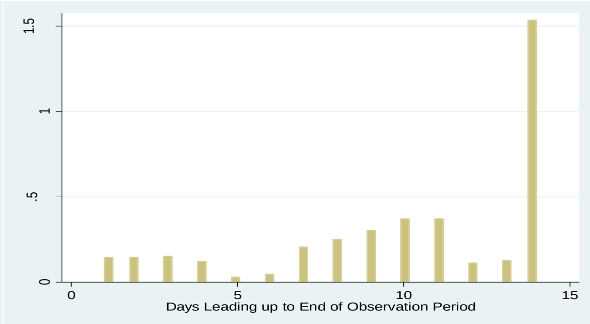
Miscellaneous Chemical



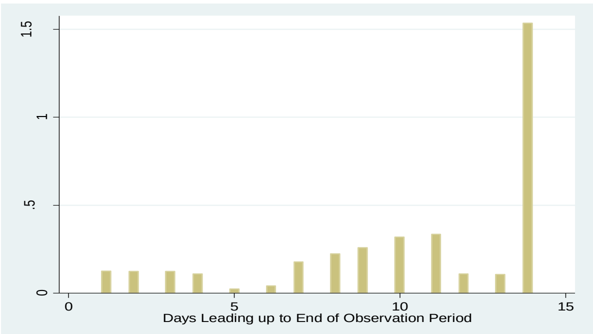
Communications



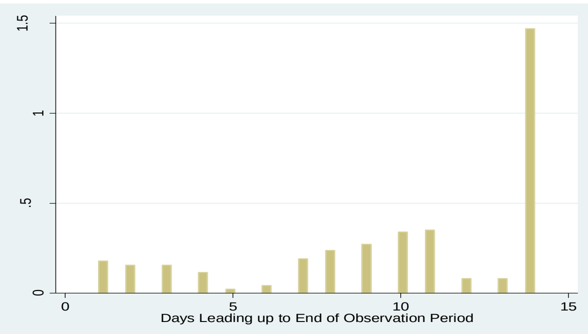
Computer Hardware & Software



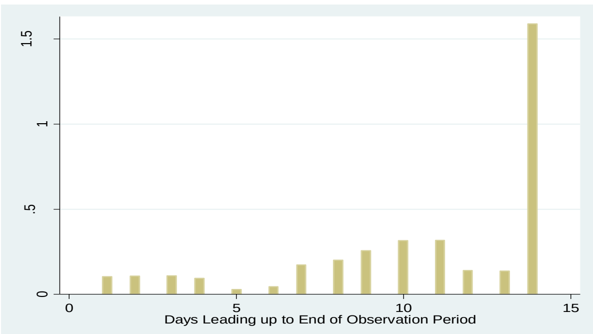
Computer Peripherals



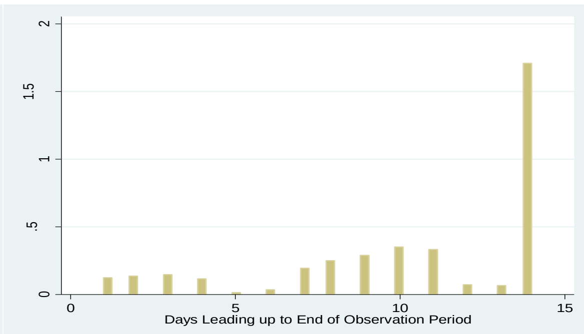
Information Storage



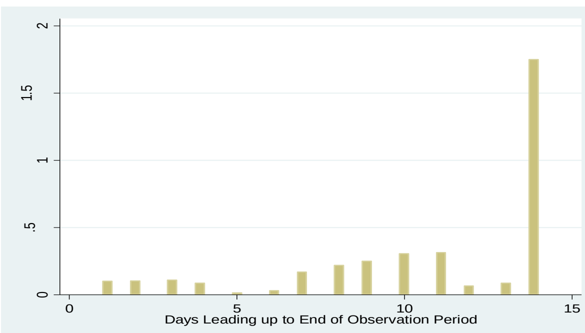
Electronic Business Methods and Software



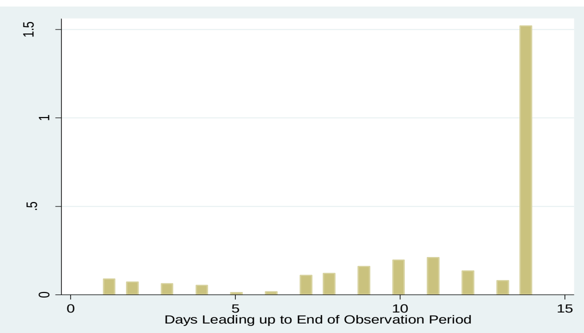
Drugs



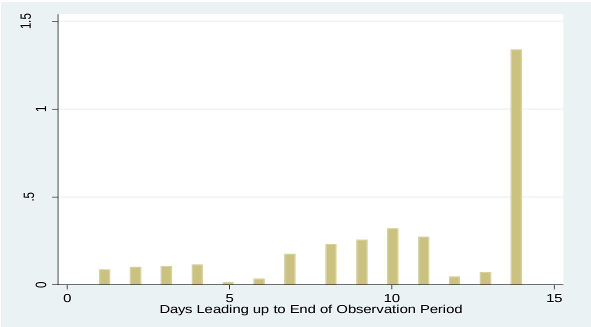
Surgery and Medical Instruments



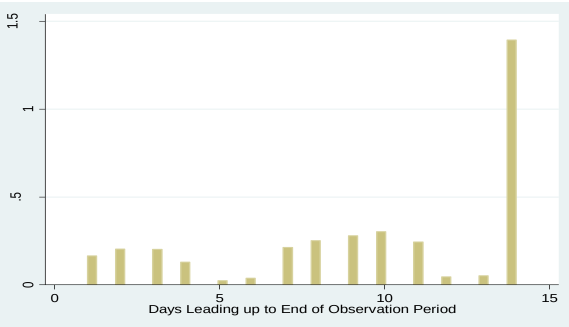
Genetics



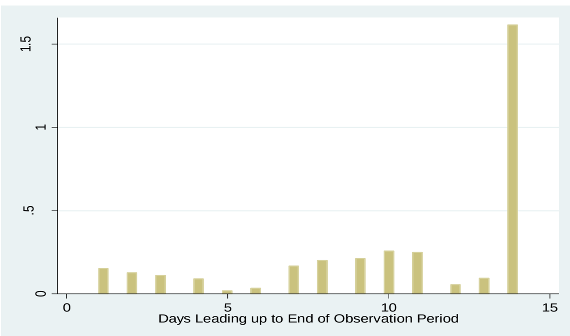
Miscellaneous Drugs and Medical



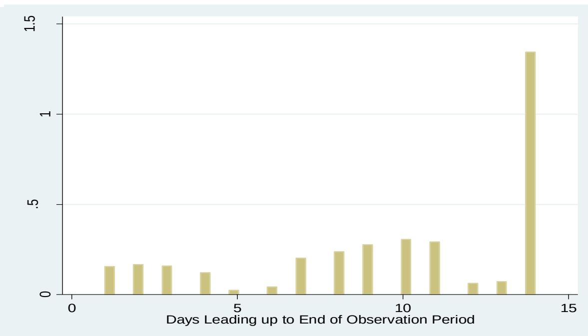
Electrical Devices



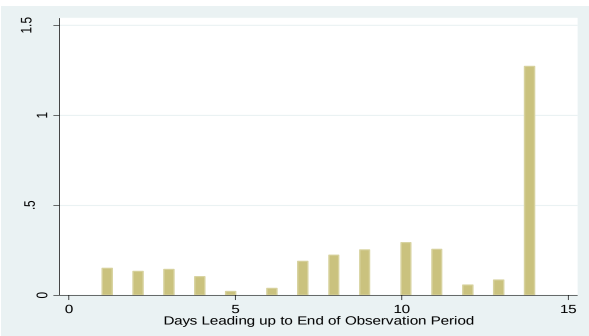
Electrical Lighting



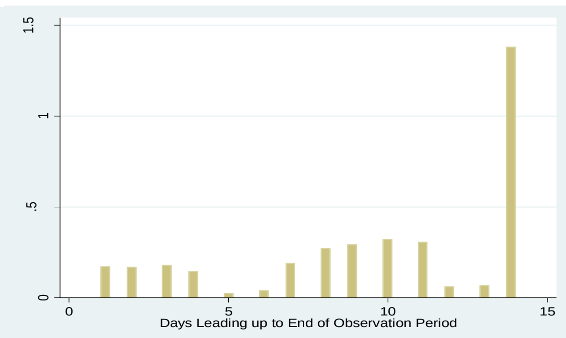
Measuring and Testing



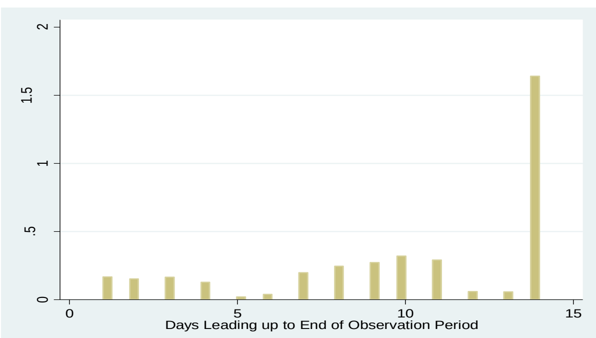
Nuclear & X-Rays



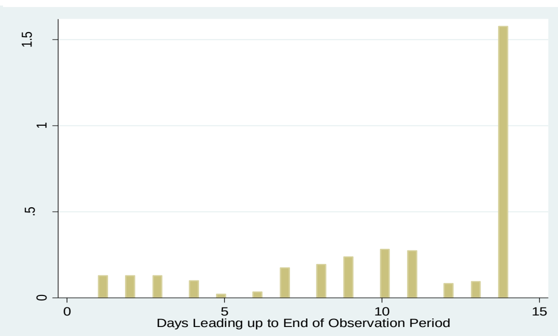
Power Systems



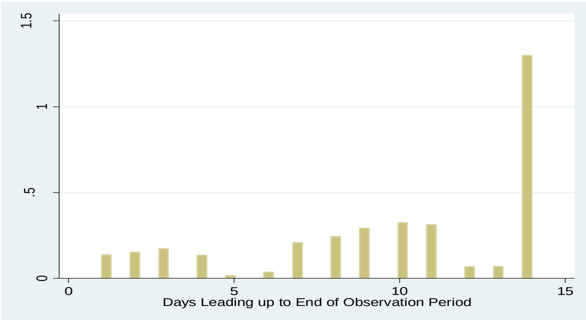
Semiconductor Devices



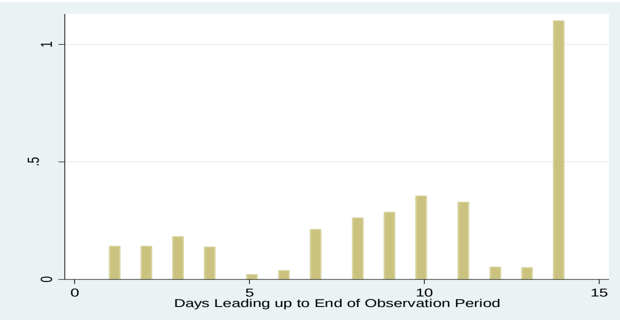
Miscellaneous Electrical



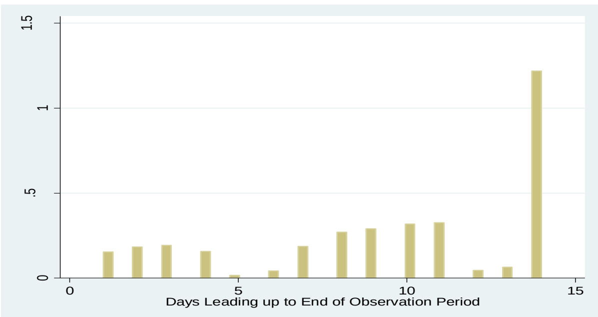
Materials Processing and Handling



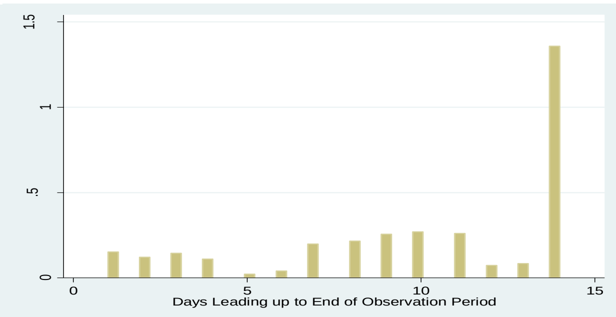
Metal Working



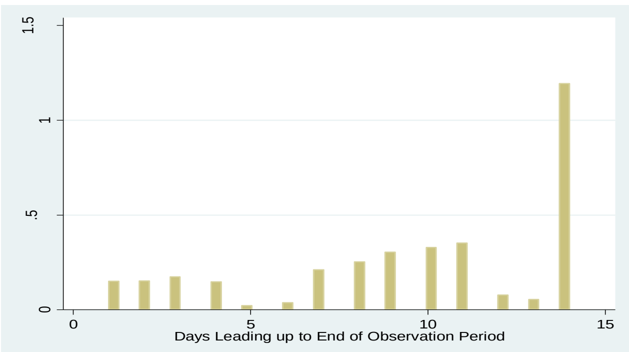
Motors, Engines and Parts



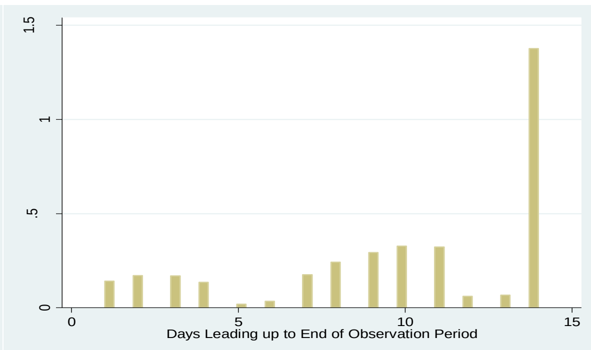
Optics



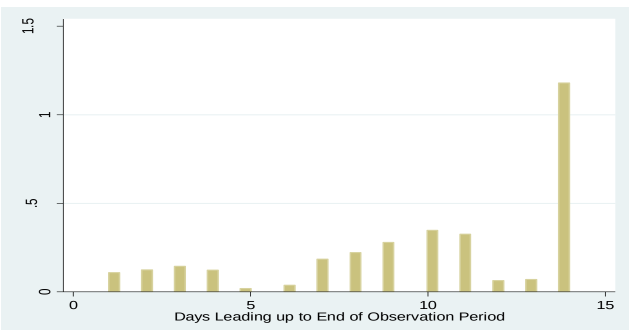
Transportation



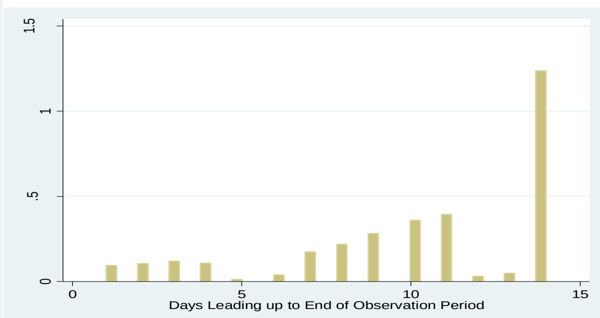
Miscellaneous Mechanical



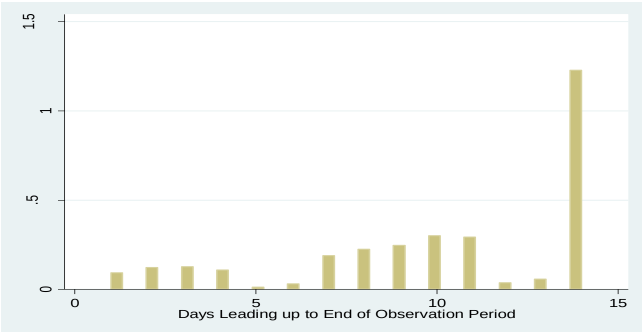
Agriculture, Husbandry and Food



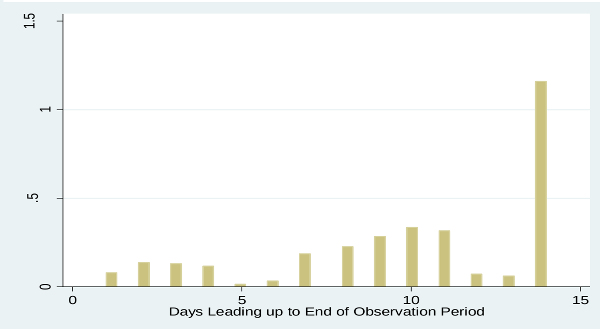
Amusement Devices



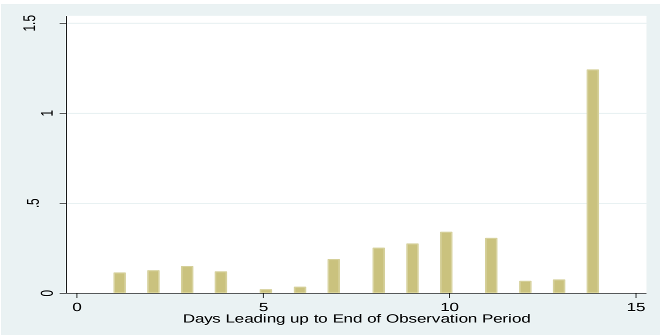
Apparel and Textiles



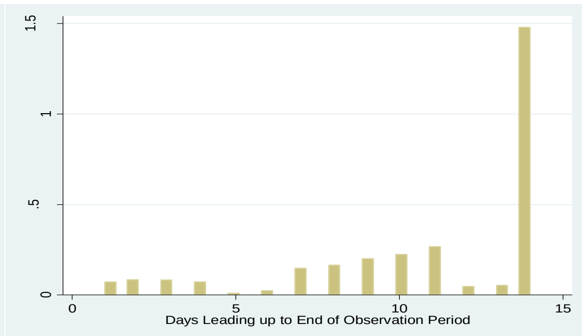
Earth Working and Wells



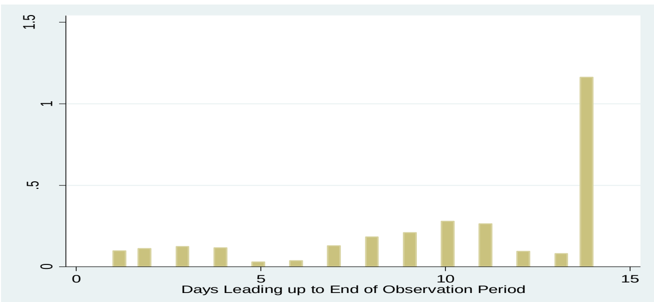
Furniture & House Fixtures



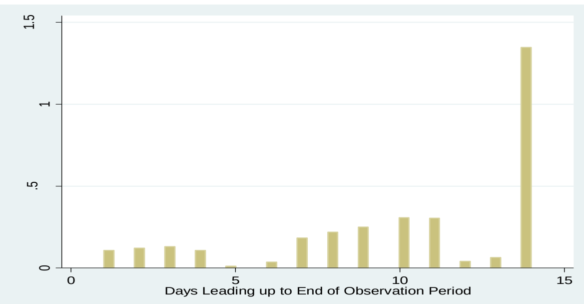
Heating



Pipe & Joints



Receptacles



Miscellaneous Other

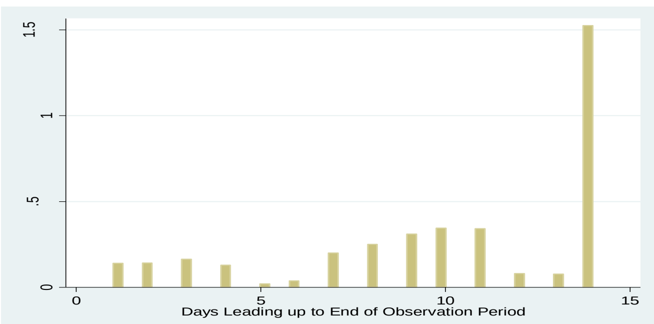


TABLE A1

Relationship between Examiner Telecommuting and the Likelihood that First Office Actions Are Completed on the Last Day of the Quota Period: Event-Study Results

	(1)	(2)
(Omitted: > 4 Years Prior to Telecommuting Dummy)		
4-Years Prior to Telecommuting Dummy	0.002 (0.004)	0.002 (0.004)
3-Years Prior to Telecommuting Dummy	0.009* (0.005)	0.008 (0.005)
2-Years Prior to Telecommuting Dummy	0.014** (0.006)	0.010* (0.006)
1-Year Prior to Telecommuting Dummy	0.010 (0.006)	0.003 (0.006)
Year Starting Telecommuting Dummy	0.072*** (0.008)	0.065*** (0.008)
1-Year Post Starting Telecommuting Dummy	0.051*** (0.009)	0.046*** (0.009)
2-Years Post Starting Telecommuting Dummy	0.021** (0.009)	0.016* (0.009)
3-Years Post Starting Telecommuting Dummy	0.017 (0.010)	0.011 (0.010)
>= 4-Years Post Starting Telecommuting Dummy	0.001 (0.012)	-0.003 (0.012)
Examiner and Year Fixed Effects?	YES	YES
Other Covariates?	NO	YES

Standard deviations are indicated in parenthesis and are clustered at the examiner level. Estimates are from a sample of 1,741,500 first office actions completed between March, 2001 and July, 2012 (focusing only on applications filed after March, 2001).

TABLE A2

Mean Rates of End-loading of First-Office Actions on the Merits at Quota-Period Ends, Separately by Examiner Status

	(2)	(3)
	APPLICATIONS REVIEWED BY THOSE WITHOUT SIGNATORY AUTHORITY	APPLICATIONS REVIEWED BY THOSE WITH SIGNATORY AUTHORITY
End-loading Rate	0.532 (0.498)	0.436 (0.495)

Standard deviations are indicated in parenthesis. Statistics are from a sample of 1,741,500 first office actions completed between March, 2001 and July, 2012 (focusing only on applications filed after March, 2001). Examiners with General-Schedule pay levels below GS-13 need supervisory approval on the first office actions that they submit, while those above GS-13 need no such approval.

TABLE A3

Falsification Tests. Relationship between Examiner Telecommuting and Certain Immutable Characteristics of the Application: Dynamic Difference-in-Difference Regression Results

	(1)	**(2)**
	Incidence of Large Entity Applicant (Mean = 0.72)	Incidence of Previous Filing at the EPO or JPO (Mean = 0.06)
(Omitted: > 4 Years Prior to Telecommuting Dummy)		
4-Years Prior to Telecommuting Dummy	-0.002 (0.002)	-0.002* (0.001)
3-Years Prior to Telecommuting Dummy	-0.000 (0.003)	-0.003** (0.001)
2-Years Prior to Telecommuting Dummy	-0.002 (0.003)	-0.003** (0.001)
1-Year Prior to Telecommuting Dummy	-0.006* (0.003)	-0.005*** (0.001)
Year Starting Telecommuting Dummy	-0.002 (0.004)	-0.004** (0.002)
1-Year Post Starting Telecommuting Dummy	0.001 (0.004)	-0.006** (0.002)
2-Years Post Starting Telecommuting Dummy	0.002 (0.005)	-0.006** (0.002)
3-Years Post Starting Telecommuting Dummy	0.004 (0.005)	-0.006** (0.003)
>= 4-Years Post Starting Telecommuting Dummy	0.005 (0.006)	-0.008** (0.003)
Examiner and Year Fixed Effects and Other Covariates?	NO	YES

Standard deviations are indicated in parenthesis and are clustered at the examiner level. Estimates are from a sample of 1,741,500 first office actions completed between March, 2001 and July, 2012 (focusing only on applications filed after March, 2001). All regressions include examiner and year effects.